SOLOMON GARTENHAUS

PHYSICS

BASIC PRINCIPLES

Without my work in natural science I should never have known human beings as they really are. In no other activity can one come so close to direct perception and clear thought or realize so fully the error of the senses, the mistakes of the intellect, the weakness and greatness of the human character.

GOETHE

SOLOMON GARTENHAUS

PHYSICS
BASIC PRINCIPLES

To Johanna

Preface

This combined edition is designed for a one-year introductory physics course for students of science and engineering. Approximately one third of this volume (Chapters 2–12) is devoted to mechanics. The remainder deals with kinetic theory, heat, and thermodynamics (Chapters 13–17), wave motion (Chapter 18), electromagnetism (Chapters 19–29) optics (Chapters 30–33), and concludes with an introductory chapter on quantum phenomena.

Although most of the subject matter of this volume deals with the laws and the phenomena of classical physics, a substantial number of references to, and illustrations of, nonclassical phenomena have been distributed throughout the text. These are intended to give the student some feeling for recent developments, as well as to alert him to the limitations of certain of the classical laws, and of the seemingly universal applicability of others. Thus, he sees the ideas of the conservation of energy and of momentum applied not only to laboratory-sized bodies, but to the motions and the interactions of nuclear, atomic, planetary, and stellar systems as well. In addition, a number of chapters contain optional sections (marked by a dagger †), which generalize or extend the text material in some way or describe related but more recent discoveries. For example, Chapter 1 includes a qualitative description of nuclei and of subnuclear particles, and in Chapter 19 the charge structure of the neutron and the proton is briefly discussed. Other material of this nature includes discussions of: the Lorentz transformation, as the high

velocity generalization of the Galilean transformation (Section †3–12); the relativistic forms of the laws of energy and momentum conservation (Section †9–10); and Einstein's equivalence principle and the bending of light rays in a gravitational field (Section †5–9).

To a considerable extent, this book reflects my view that the student for whom the book is intended is preparing for a career for which some in-depth knowledge of physics is essential. With today's accelerating tempo of scientific discovery and technological innovation foremost in mind, I have endeavored, throughout the text, to emphasize the basic physical laws and to present them in as succinct and complete a manner as it is possible to do at the introductory level. Included therefore are not only extensive discussions of the scope of the laws and their interrelations but, where appropriate, their ranges of validity as well. By placing the focus on the basic physical laws in this way, it is hoped that the student will come to understand some of the physicists' empirical-logical approach to the world and will acquire thereby a realistic perception of these laws and a secure intellectual base from which to pursue further studies.

Specific features of the book which are directed toward implementing this approach include:

1. All basic physical laws, such as Newton's laws of motion, Coulomb's law, and Faraday's law, are first introduced by use of certain conceptually simple and easily visualized experiments. The laws themselves are then abstracted from these experiments and reformulated in quantitative mathematical terms. Finally, the important features of the laws and their relation with other laws are brought out and rëinforced in the student's mind by applying them to a variety of physical situations.
2. Because of their conceptual importance, both as a starting point in physical reasoning and as a tool for codifying diverse physical phenomena, the various conservation laws receive particularly heavy emphasis. More than a third of the material on mechanics, for example, deals with these laws. Extensive discussions of these laws are presented and the student is able to see, at several levels, their connection with other physical laws, their interrelations, and their limits of validity.
3. Wherever appropriate, all physical quantities are first introduced by use of operational definitions. My view here is that in the student's subsequent capacity as a professional he may well be called upon to apply physical laws in regions bordering on their limits of validity. Therefore, he should be able to reexamine critically and systematically all basic definitions and physical laws in light of their underlying assumptions and for this purpose operational definitions are indispensable.
4. A considerable effort has been made to present the material in a way to make it dovetail more closely with that found in more advanced courses. Thus, the language, the basic approach, and the attitude, but not the level, are those found more often in texts at the intermediate and higher level.

Some other features of the book are:

1. As is traditional in introductory physics courses of this type, it is assumed that the student is taking concurrently a course in calculus. Nevertheless, many of the basic ideas of the differential and the integral calculus are introduced as needed and integrated with the text. Detailed derivations, however, are usually relegated to one of the brief mathematical appendices. Vectors are introduced in Chapter 3 and used throughout the text. The dot product is defined in Chapter 7 and the cross product in Chapter 10.
2. Each chapter contains a wide selection of *questions* and *problems*. The more difficult problems have been marked by an asterisk (*) and those which require some knowledge of material covered in optional sections have been similarly marked by a dagger (†). As a general rule, the order of the problems follows closely the ordering of the text material. The questions generally call for answers of a qualitative or semiquantitative nature, and are intended mainly to help the student test his physical understanding of the text material. Also included are some questions whose main purpose is to be thought provoking.
3. A sizable number of worked examples (averaging more than ten per chapter) are distributed throughout the text. These span all levels of difficulty, from the simple formula-substitution type to the more complex ones whose purpose is to extend some aspect of a topic treated in the text. Also, many of these examples serve as models which the student can use as an aid in problem solving.
4. The SI or International System of Units (often called the metric or the MKSA system) is used exclusively throughout the text. Included, however, are definitions of and conversion tables to the CGS and the English system of units. With only rare exception the answers to all problems and worked examples that call for a numerical value are in metric units.
5. A broad spectrum of applications has been included in the text. Thus there is available considerable latitude in adapting it to courses of varying length or specialized needs.

In writing this book, I have received help in many forms and from many more sources than it is possible to acknowledge individually. I should like to thank particularly Orland E. Johnson who has been a valuable source of advice on all aspects of the book throughout the long years of writing. I am also greatly indebted to David J. Ennis and Nicholas J. Giordano, who with skill and patience worked out solutions to all problems and were helpful in a variety of other ways; to John J. Brehm, Gary S. Kovener, Earl W. McDaniel and Walter W. Wilson, who critically reviewed the manuscript and made many valuable suggestions; to the many of my colleagues at Purdue, particularly Edward Akely, Irving Geib, Don Schleuter, and Isadore Walerstein for much encouragement and for help when needed; and to the members of the editorial staff at Holt, Rinehart and

Winston, Inc., through whose patience, persistence, and labors this book ultimately came into being. Finally, I should like to thank Debbie Carr, Coleen Flanagan, and my wife Johanna for their invaluable secretarial assistance and my sons Mike and Kevin for their help with the galleys.

Solomon Gartenhaus
Purdue University

Dec. 1976

Contents

2 Kinematics in one dimension 25

3 Two-dimensional kinematics 53

4 The laws of Newton 89

Appendices

List of tables

1 Physical quantities and their measurement

Here men from the planet earth first set foot upon the moon.
PLAQUE LEFT ON THE MOON BY THE APOLLO 11 ASTRONAUTS

1-1 General introduction

The decades since World War II have been marked by a number of major technological innovations and breakthroughs. Very noticeable among these are the following developments:

1. Controlled and uncontrolled nuclear energy sources.
2. High-speed electronic computers, and their general acceptance throughout most of society.
3. Space propulsion systems and the landing of man on the moon.

These developments are so new that even today it is difficult to assess their long-range significance. Nevertheless, it is apparent that their introduction marks a precipitous rise not only in man's ability to acquire and to process new knowledge but in his self-confidence at its acquisition as well. Certainly they have irreversibly altered the vision he has of himself and of his role in the physical world.

Very often it seems as if new technologies, such as those above, are ushered in abruptly, by some particular event. Thus we often associate the

"atomic age" with the instant, 5:30 A.M. on July 16, 1945, when the first atomic bomb was exploded at Alamogordo, New Mexico. The "space age" is similarly associated in the minds of many people with the date, July 20, 1969, when the Apollo 11 astronauts first landed on the moon. To some, the announcement of June 25, 1948, of the discovery of the transistor and its implied promise of miniature electronics represents the true beginning of the "computer age." In reality, events such as these do not in themselves mark the beginning of new technologies at all; rather they mark a high point of a period of research and discovery that generally began long before the event itself. The landing of the Apollo 11 astronauts on the moon, for example, is in this sense, the outcome of research that has been going on steadily since the times of Galileo and Newton in the seventeenth century. Some of it went on even earlier. The people involved in this process of research and discovery also span many centuries and represent many countries of the world. In the last analysis, it is only such a large-scale pooling of knowledge, in both space and time, that makes possible at all the technological developments we have witnessed during the past few decades.

Now basic to many of the scientific and engineering disciplines that have played a significant role in bringing about these and other technologies are certain underlying physical concepts and laws. The role played by these laws, or *laws of physics*, as they are known, is to systematize and to correlate the vast body of empirical knowledge that man has acquired during the preceding centuries. Without knowledge of these laws it would be inconceivable to carry out the detailed planning and the large-scale resource allocation required to generate and to nurture new and complex technologies. The science that is concerned with the discovery, and the development, of these laws is known as *physics*. It is the subject matter of this text. By its nature, physics is a very fundamental science and it interfaces with a variety of other disciplines. For many of these, some knowledge of physics is indispensable.

1-2 The major subdivisions of physics

Mainly for historical reasons, it is convenient today to divide the subject matter of physics into two classes: (a) *classical* and (b) *modern* or *nonclassical*. Qualitatively speaking, classical physics is concerned with phenomena and laws known up to the end of the nineteenth century, whereas modern physics deals with discoveries made since then. In this text we are concerned almost exclusively with the phenomena and laws of classical physics. Nevertheless, to be able to pinpoint some of the limitations of the classical laws we shall from time to time make reference to particular nonclassical phenomena.

Classical physics is further subdivided into a number of—what were originally thought to be autonomous—branches or disciplines. These are *mechanics*, *electromagnetism*, *optics*, *acoustics*, and *thermodynamics*.

Briefly stated, the discipline of mechanics, which is also basic to the other branches, deals with the study of motion and with those physical effects that can influence motion. Also of central importance to physics is the discipline of electromagnetism, which is concerned with the study of electric and magnetic phenomena and with the relations between them. The important role played by these two disciplines of mechanics and electromagnetism may be appreciated from the fact that a knowledge of some facet of one or the other is indispensable for an understanding of most phenomena of the physical world. They are truly the *sine qua non* of physics. The third branch, *optics*, deals with physical effects associated with visible light and the fourth, *acoustics*, is concerned with the study of physical effects associated with audible sound. The last, *thermodynamics*, may be characterized as that branch of physics concerned with the generation, transport, and dissipation of heat as a form of energy. Although originally each of these five disciplines was developed independently of the others, they are very much linked together via mechanics and electromagnetism.

The era of modern physics is generally marked as having begun toward the end of the nineteenth century, with the discovery of a number of physical phenomena that were in conflict with one or more basic concepts of classical physics. Just how the required modifications of classical ideas were discovered and ultimately resolved makes one of the most interesting chapters in the history of science. It is briefly discussed in Chapter 34. Basically, these conceptual alterations were of two types. One type established an upper limit to the velocities of particles for which the laws of classical physics are applicable; this is associated with Einstein's *theory of relativity*. The second may be thought of as setting a lower limit to the mass and to the linear dimensions of physical systems for which the classical laws are valid; this is associated with the theory of *quantum mechanics*. To understand these two modern theories and the phenomena they deal with, it is necessary first to study the laws of classical physics. And it is to the study of these laws that we shall now address ourselves.

1-3 Physical quantities

A basic assumption underlying all research in physics is that natural phenomena take place in accordance with certain quantitative rules. These rules are known as *physical laws* or *laws of physics*, and the ultimate goal of all research in physics is to uncover the nature and the detailed structure of these laws. Once a certain rule has been tentatively established as a candidate for a physical law, it is usually subjected to experimental testing under the most diverse conditions possible in order to ascertain whether or not it is indeed a physical law and, if so, what the limits of its validity are. Only after it has withstood extensive testing over long periods of time do we elevate it to the status of a *law of physics*.

Based on results accumulated over several hundred years, we know

that physical laws can be expressed in the form of mathematical relations involving certain entities—such as time, or distance, or speed, or temperature—each of which is itself subject to experimental observation. We call these entities *physical quantities.* Associated with each physical quantity there is in general both a range of numerical values and a *unit.* For a particular physical situation, the numerical value associated with any given physical quantity tells us its value in terms of the unit associated with it. Thus when we say a rod has a length of 11 inches, we mean that it is 11 times longer than is a "unit rod" of length 1 inch. For the case of an automobile traveling at a speed of, say, 60 "miles per hour," the numerical value associated with the physical quantity of speed is in this case 60, in units of "miles per hour." This means that the automobile travels 60 times faster than does one that goes at a unit speed of 1 mile per hour. Note that even after the units in terms of which a given physical quantity is to be measured have been specified, the physical quantity itself can take on a variety of numerical values depending on the prevailing physical situation. It is one of the main functions of the laws of physics to tell us which particular value, out of some allowable set, a physical quantity will assume under given experimental conditions.

Now in order that the laws of physics be useful as quantitative predictive tools, it is vital that all physical quantities that enter into these laws be precisely defined. For otherwise there would be little point in having physical laws in the form of precise mathematical relationships. We say in this connection that a physical quantity has been defined in *operational terms*, or has been given an *operational meaning*, provided it has been described in terms of a set of reproducible laboratory operations that can be used to measure a unique value for the physical quantity. Thus in order to give an operational meaning to the physical quantity of length, or of time, it is necessary to describe how the length of any object, or the time interval between any two events, can be measured.

Consider for example, the three statements:

1. That stick is 2 feet long.
2. There are 608 pages in this book.
3. The temperature of that man is 98.6 degrees Fahrenheit.

The operational meaning of these statements is readily apparent to most people. Thus the first one tells us that the stick has the same length as do two 1-foot rulers placed end to end. Similarly the second and the third derive precise meanings from our commonly shared understanding of the process of counting and of taking readings on a thermometer. Let us contrast these statements now with the following:

1. The radius of the hydrogen atom is 5.3×10^{-11} meter.
2. The mass of the earth is about 6.0×10^{24} kilograms.
3. The number of oxygen molecules in this room is 8.2×10^{24}.

Even granting the availability of operational definitions for the units of the meter and the kilogram, except among people knowledgeable in these matters there is no commonly accepted understanding of what these and similar statements mean. Certainly, the first statement cannot be confirmed by use of an ordinary meter stick nor the second by using a laboratory scale. Nor is it conceivable to count the number of oxygen molecules in a room in the same way that we count the number of pages in a book. To give a meaning to statements of this type, operational definitions are essential. What is required, in other words, is a practical procedure—not one that can be carried out only in principle—by means of which all of the necessary measurements can be made. Only after an operational meaning has been given to such statements can use be made of the values of the physical quantities that they imply.

As an illustration of this matter, consider the statement that the circumference of the earth is about 40,000 kilometers. What operational meaning can we give this statement? Eratosthenes (circa 250 B.C.), the ancient Greek philosopher, pondered this very problem of the size of the earth and many years ago gave it an operational meaning by devising a very ingenious method of measuring it. He noted that on the first day of summer in the city of Syene, Egypt, the sun was vertically overhead at noon. Specifically, he observed that at this instant sunlight would appear at the bottom of a very deep vertical well, thus implying that a line drawn from the center of the earth to the sun went through Syene. Moreover, he knew that at noon on the same day in the city of Alexandria, which was about 5000 *stadia* (the *stadium* was a Greek unit of length which is today believed to be about 0.16 kilometer) due north of Syene, the sun would make an angle of about 7° with the vertical. Reference to Figure 1-1 shows that since there are 360° in a circle, the circumference of the earth may be computed from these data to be $(360 \times 1/7) \times 5000$ stadia $= 2.6 \times 10^5$ stadia. In this way, therefore, Eratosthenes not only gave an operational meaning to the radius of the earth, but also found a value that today is believed to have been within 20 percent of the presently accepted one.

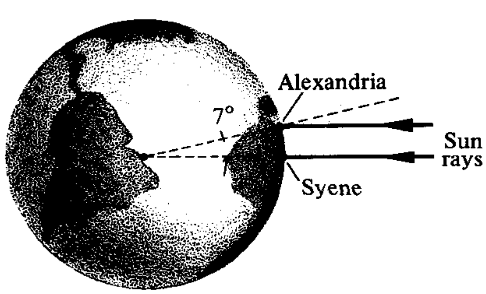


Figure 1-1

1-4 Units and standards

A properly defined physical quantity will have associated with it a set of explicit instructions by means of which its value for any given physical situation can be measured. In general, the result of such a measurement is a numerical value expressed in some system of units. We shall now define what is meant by the term *units* and the related one of *standard.*

The *unit* of a *physical quantity* is a *value or a magnitude in terms of which other values or magnitudes of the physical quantity may be expressed.* Thus a speed of 15 meters/second represents one that is 15 times that of a "unit" speed of 1 meter/second, and a mass of 10^{-3} kilogram represents one that is one thousandth that of a unit mass of a kilogram. Related to this notion of a unit of a physical quantity is that of the *standard for a unit.* A *standard* is the *physical embodiment of a unit.* That is, a standard is a particular physical system maintained under specified external conditions and with a unique physical property that defines the given unit. By contrast to a *unit,* which is fixed by definition and thus is independent of external conditions such as temperature or pressure, a standard is generally dependent on the physical environment. As a rule, a given standard is the appropriate physical realization of a unit only under certain restricted, but realizable, external conditions.

Consider in Figure 1-2a two very thin and straight rods A and B and imagine carrying out a relative measurement of the lengths of the two rods with the result that B is, say, 2.5 times as long as A. In operational terms this means that if, as in Figure 1-2b, we take three straight rods, two of them with the same length as A and one with a length one half that of A, and lay them end to end in a straight line, then the resultant length is the same as that of B. Suppose now that A is adopted as the standard used to define a unit of length. Then B will have a length of 2.5 of these units. Furthermore, this standard can be used to express the length of any other rod in terms of it and, in this way, provided that a replica of rod A is universally available, a standard for a unit of length has been defined in operational terms. Note that this definition for the standard makes sense only for purposes of measuring lengths that are comparable to that of A. To measure lengths that exceed that of A, say by a factor of 10^5, or are less than it by a factor of 10^{-5}, a more detailed and operational description of the necessary measuring process is required.

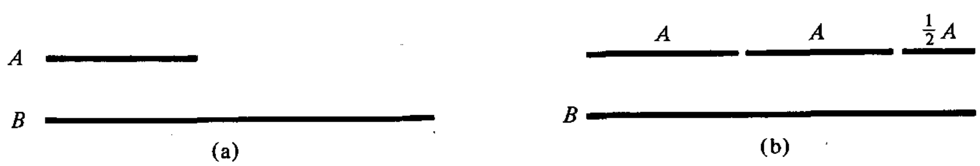


Figure 1-2

1-5 The International System of units

The problem of defining units and the associated one of devising an acceptable standard goes back to antiquity, and in the course of time many systems of weights and measures have been devised. In addition to the Greek unit of length of the *stadium*, another important unit of many ancient peoples was the *cubit*, which was defined to be the distance from a man's elbow to the end of his middle finger. Similarly, the *inch* was originally defined by the Romans to be a "thumb's breadth," and a *mile* as 1000 double steps. Up to the end of the eighteenth century there existed no generally accepted system of weights and measures, and each country had its own system. Of particular interest to us is the system of weights and measures known as the *English system*. Its basic units are the *pound*, the *yard*, and the *second* and it remains in usage today only in some of the Commonwealth countries and the United States. Except for the English speaking people, who had this common system of weights and measures, up to the time of the French Revolution there existed no generally accepted system of weights and measures among other national groups.

In May 1790, in an effort to create some uniformity out of this chaos, the National Assembly of France met and called upon the French Academy of Sciences to join with the Royal Society of London in order to "deduce an invariable standard for all of the measures and all of the weights." At this time the English had already available an acceptable system and thus declined the invitation. The French Academy thereupon decided to tackle the problem alone and developed the system that today we know as the *metric system*. Its unique and distinguishing feature is that all multiples and submultiples of the basic units in this system are related to each other by simple factors of ten. Two units basic to this system are defined below and are the unit of mass of the *kilogram* and the unit of length of the *meter*. The word "meter" itself comes from the Greek *metron*, which means "a measure." Originally it was intended that the meter should represent one ten-millionth (10^{-7}) of the distance from the north pole to the equator as measured along the meridian running near Barcelona, Dunkirk, and Paris. This definition is, however, no longer in use today.

In 1872 representatives of 26 countries met in France and drew up a treaty, called the *Metric Convention*, which was signed in 1875 by 17 countries. The United States was one of these signatories. The treaty established metric standards for length and mass, created the BIPM (Bureau International des Poids et Mesures), and called for a CGPM (Conférence Générale des Poids et Mesures), to meet every six years to consider possible improvements in standards. Finally, the treaty created the CIPM (Comité International des Poids et Mesures), whose function was to implement the recommendations of the CGPM and to direct the activities of the BIPM.

Since its creation almost 100 years ago, this international structure has functioned smoothly and is recognized today as the undisputed and ultimate authority on all questions relating to weights and measures.

In 1960 the CGPM initiated the establishment of our present-day metric system, which is called the *International System of Units*, or *SI* for short. At this conference our present standard for the meter was established, and this was followed at the 1967 meeting with the establishment of the modern standard for (a) the unit of time of the *second*; (b) the unit of temperature of the *kelvin*; and (c) the unit of luminous intensity of the *candella*. Together with the standard for the unit of mass of the *kilogram*, which was established at the CGPM in 1901, and the unit of electric current of the *ampere*, which was established at the CGPM in 1946, these six fundamental units constitute the basic SI units. All other units are deriveable from these. In our studies we shall be using SI units almost exclusively.

Since its creation nearly two centuries ago, the number of countries utilizing the metric system has grown enormously. Among the technologically advanced nations, only the United States and a few Commonwealth countries still utilize the English system of units. However, there are indications that by the twenty-first century the metric system will dominate completely. At the time of this writing (1973), England is in the process of changing over, and there are indications that we in the United States shall follow soon. In 1968, for example, Congress authorized the Secretary of Commerce to study the feasibility of our converting to the metric system, and since 1970 the National Aeronautics and Space Administration (NASA) has required all technical reports, notes, memoranda, and so on, to make use of SI units exclusively.

1-6 The SI mass standard

In this and the next two sections we shall define the standards for the three basic SI units of mass, length, and time. As we shall see later, all other mechanical units such as force, energy, and power can be expressed directly in terms of these three.

The *mass* of a given body has to do with the amount of matter present in it. In microscopic terms, this means that the mass of any system is a measure of the total number of atoms and molecules of that system. Qualitatively speaking, the more massive is a macroscopic system, the heavier it will appear to be.

At the third meeting of the General Conference of Weights and Measures (CGPM) in 1901 the *kilogram* (kg) was defined to be equal to the "mass of the international prototype of the kilogram," which is a cylinder of platinum iridium alloy kept by the International Bureau of Weights and Measures (BIPM) at Sèvres, France. There has been no change in the definition of this mass standard since the meeting in 1901. A duplicate of this standard is kept

at the National Bureau of Standards in Washington, D.C., and is known as *Prototype Kilogram No. 20* (see Figure 1-3). It is a platinum iridium cylinder, 3.9 centimeters in diameter and 3.9 centimeters in height, and serves as the mass standard for the United States.

The gram (g) is a unit of mass defined to be one thousandth of a kilogram. Thus

$$1 \text{ g} = 10^{-3} \text{ kg}$$

or, equivalently,

$$1 \text{ kg} = 10^{3} \text{ g}$$

In the English system the basic unit of mass is the *slug*. As described in more detail in Chapter 4, today this unit is defined in terms of the standard for the kilogram. The problem of how to determine the mass of an arbitrary body in terms of this standard will also be considered there.

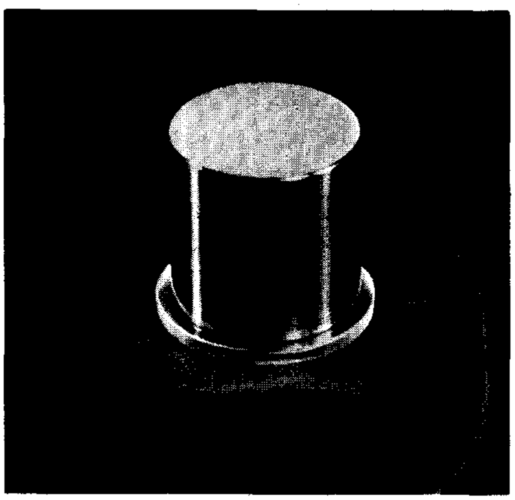


Figure 1-3 The Prototype Kilogram No. 20, which is housed at the National Bureau of Standards. It is a platinum, iridium cylinder 39 mm in diameter and 39 mm high and is an accurate copy of the international standard kept at the BIPM in Sèvres, France. It defines a mass standard accurate to one part in 10^8. (Courtesy U.S. Bureau of Standards.)

1-7 The SI standard for length

At the 1960 meeting of the CGPM the presently accepted standard for the meter was adopted. Taking advantage of the extraordinarily accurate measurements of length that have been made possible by the development of precision interferometers, the standard for the meter was defined at this conference to be:

> ... a length equal to 1,650,763.73 wavelengths in vacuum of the radiation corresponding to the transition between the levels $2p_{10}$ and $5d_5$ of the krypton-86 atom

and this standard is universally accepted today. It replaces the one in usage prior to 1960 in which the meter was defined in terms of the distance between two lines on a certain platinum bar maintained under specified conditions. Two of the advantages of this modern standard over the previous ones are:

1. It is readily available to anyone who has access to a well-equipped laboratory.
2. The standard is much more precisely defined—to nine significant figures—than any of its precursors.

In addition to the meter it is convenient to define certain multiples and submultiples of this unit. Table 1-1 lists a number of prefixes associated with particular multiples and submultiples of the meter. According to the table, for example, a centimeter is abbreviated "cm" and represents 10^{-2} meter and a kilometer is abbreviated "km" and represents 10^3 meters. Similarly, "Tm" is the symbol for a terameter and represents a distance of 10^{12} meters and "mm" is the symbol for a millimeter, which is 10^{-3} meter. The prefixes and their abbreviations as listed in Table 1-1 apply not only to the unit of length but also to all SI units. Thus, as we saw in the previous section, $1\ \text{kg} = 10^3$ grams. Similarly, $1\ \mu\text{g} = 10^{-6}$ gram = 1 microgram and $1\ \text{Gg} = 10^9$ grams.

Table 1-1 Some prefixes for multiples and submultiples of all SI units

Multiple	*Prefix*	*Symbol*
10^{12}	tera	T
10^9	giga	G
10^6	mega	M
10^3	kilo	k
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p

The unit of length in the English system is today *defined* in terms of the SI standard of length. The precise definition is

$$1\ \text{yd} = 0.9144\ \text{m} \quad \text{(exactly)}$$

or, equivalently,

$$1\ \text{in} = \frac{1}{36}\ \text{yd} = 2.54\ \text{cm} \quad \text{(exactly)}$$

Roughly speaking, 1 ft $\left(=\frac{1}{3}\ \text{yd}\right)$ is approximately 30 cm, or to 10 percent accuracy, a meter is approximately 1 yd.

1-8 The SI standard for time

The third basic unit needed in a study of physics is that of *time*. In this section we shall define the standard for the second (s), which is the SI unit for time.

One of the earliest time standards is *universal time*. Basic to this standard is the "day," which is defined as the time interval at a fixed point on earth between noon on one day to noon on the following day. The *mean solar day* is then defined as the length of this day when averaged throughout the year. The basic unit of universal time is the *mean solar second*, which represents 1/86,400 of the mean solar day.

Unfortunately, the mean solar second is *not* acceptable as a standard. The reason for this has to do with the fact that as a consequence of frictional effects associated with lunar and solar tides, the earth's period of rotation about its axis is increasing. Experiment shows that, in effect, the length of the day is gradually increasing at the very slow but nevertheless perceptible rate of 10^{-3} second per century. With this in mind, at the 1956 CGPM, the second was redefined to be exactly 1/31,556,925.9747 part of the tropical year 1900. Time, as defined in this way in terms of the orbital motion of the earth, is called *ephemeris time* and this unit of time itself is known as the *ephemeris second.*

At the 1967 General Conference on Weights and Measures, the *second* was redefined once again. Just as for the new standard for the meter, which had been established in 1960, by the year 1967 technological advances had made it possible to define a time standard that is much more accurate and conveniently measurable than is the ephemeris second. Figure 1-4 shows a device known as an *atomic clock*, which can be used to measure time intervals with extraordinary accuracy. Underlying the operation of such a clock is the fact that the resonant frequencies (the reciprocals of time intervals) associated with certain atomic transitions or vibrations are relatively insensitive to changes in their environment. Moreover, these frequencies themselves can be measured to very high precision. Accordingly, at the 1967 meeting of the CGPM the second was redefined as:

> . . . the duration of 9,192,631,770 cycles corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom.

This standard has the virtue that it is readily accessible everywhere in the world and is not subject to variations due to any known causes. Today, it is universally accepted in science, industry, and commerce.

As noted before, the prefixes defined in Table 1-1 apply equally to all SI units including that of the *second*. Thus using the symbol "s" for the second, we have, for example, $1\ \mu s = 1$ microsecond $= 10^{-6}$ second and 1 ns $=$ 1 nanosecond $= 10^{-9}$ second.

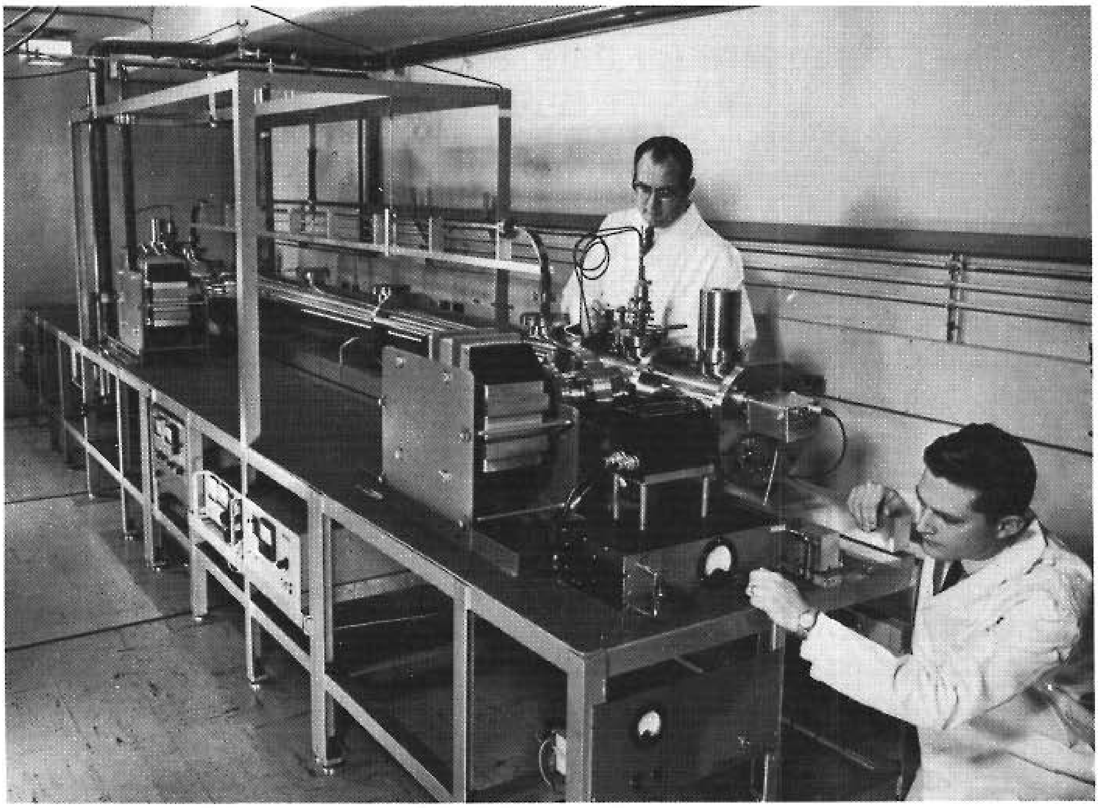


Figure 1-4 *Scientists at the laboratories of the National Bureau of Standards in Boulder, Colorado, are shown operating a cesium-133 atomic clock. This type of clock makes it possible to define a standard for time accurate to one part in 10^{12}. (Courtesy U.S. Bureau of Standards.)*

1-9 Conversion of units

In working problems whose answers involve numerical values for physical quantities, it is frequently necessary to convert a numerical value from one system of units to another. A typical case might call for converting a speed of, say, 3 miles per hour into the SI unit for speed, of meters per second. Another might involve expressing the radius of the earth, which is 6.4×10^6 meters, in terms of megameters (Mm). In carrying out such conversions of units it is generally advisable to adopt the procedure of treating the units for physical quantities as ordinary algebraic quantities, which are themselves subject to the rules of algebra. For example, to convert a time interval of 320 seconds into minutes (min) we first note that there are 60 seconds in 1 min, and thus the ratio 1 min/60 s has the value unity. Hence,

$$320\text{ s} = 320\text{ s} \times \frac{1\text{ min}}{60\text{ s}} = \frac{320}{60}\text{ min} = 5.3\text{ min}$$

where the units of seconds (s) have been canceled from the numerator and denominator. Similarly, to convert a speed of $5.0\text{ mm}/\mu\text{s}$ to meters per

second (m/s), we note that according to Table 1-1 the quantities 10^{-3} m/mm and 1 μs/10^{-6} second both have the value unity. Hence

$$\frac{5.0\text{ mm}}{\mu\text{s}} = \frac{5.0\text{ mm}}{\mu\text{s}} \times \frac{10^{-3}\text{ m}}{\text{mm}} \times \frac{1\ \mu\text{s}}{10^{-6}\text{ s}} = 5.0 \times 10^{3}\text{ m/s}$$

where the units of μs and of mm have been canceled in the numerator and in the denominator.

Example 1-1 The mass of the earth is 6.0×10^{27} grams. Express this mass in units of the kilogram, the milligram, and the teragram.

Solution According to Table 1-1 the following three ratios:

$$\frac{10^{-3}\text{ kg}}{1\text{ g}} \qquad \frac{10^{3}\text{ mg}}{1\text{ g}} \qquad \frac{10^{-12}\text{ Tg}}{1\text{ g}}$$

all have the value unity. Hence

$$6.0 \times 10^{27}\text{ g} = 6.0 \times 10^{27}\text{ g} \times \frac{10^{-3}\text{ kg}}{\text{g}} = 6.0 \times 10^{24}\text{ kg}$$

In the same way we obtain the equivalent values for the mass of the earth of 6.0×10^{30} mg and 6.0×10^{15} Tg.

Example 1-2 The density—that is, the mass per unit volume—of water at room temperature is 1.0 g/cm^3. What is the density of water in terms of the unit of kg/m^3?

Solution Making use of Table 1-1, we have

$$1 = \frac{1\text{ kg}}{10^{3}\text{ g}} = \frac{10^{2}\text{ cm}}{1\text{ m}}$$

from which

$$\frac{1.0\text{ g}}{\text{cm}^3} = \frac{1.0\text{ g}}{\text{cm}^3} \times \frac{1\text{ kg}}{10^{3}\text{ g}} \times \frac{(10^{2}\text{ cm})^3}{\text{m}^3} = 1.0 \times 10^{3}\text{ kg/m}^3$$

Example 1-3 What is the speed in yards per second and in meters per second of an automobile traveling at 60 miles per hour (mi/hr)?

Solution Since there are 1760 yd in a mile and 3600 seconds in an hour, it follows that

$$\frac{60\text{ mi}}{\text{hr}} = \frac{60\text{ mi}}{\text{hr}} \times \frac{1760\text{ yd}}{\text{mi}} \times \frac{1\text{ hr}}{3600\text{ s}} = 29\text{ yd/s}$$

Moreover, since by definition 1 yd = 0.9144 meter, we find similarly that

$$29\,\frac{\text{yd}}{\text{s}} = 29\,\frac{\text{yd}}{\text{s}} \times \frac{0.9144\text{ m}}{\text{yd}} = 27\text{ m/s}$$

1-10 The chemical elements

In studies of physics it is often very helpful to be able to describe the behavior of macroscopic systems in terms of the properties of their constituent atoms and molecules. The purpose of this section is to review very briefly the properties of atoms and to gather together certain important results for future reference. Much of this material is normally treated in introductory studies of chemistry.

Although the notion that matter consists of certain indivisible entities or *atoms* goes back to the Greek philosophers, particularly Democritus (circa 450 B.C.), our present-day view of the atomic nature of matter has evolved mainly during the past century. The concept itself is known as the *atomic hypothesis*, and is one of the most important cornerstones of modern science. It is universally accepted today.

According to this hypothesis, for each substance there is a smallest subdivision that has all the essential properties of that substance and whose further breakdown leads to components with different properties. For the case of the elements, such as helium or zinc or uranium, these smallest subdivisions are called *atoms*; for chemical compounds, such as water or carbon dioxide, they are known as *molecules.* A molecule is itself composed of two or more atoms. For example, a water (H_2O) molecule consists of two hydrogen atoms and one oxygen atom held together by their chemical binding forces. We know today that there exist at least 92 distinct types of stable atoms and that all chemical substances that we know of are made up of these. In addition, during the last two decades a number of new but unstable elements—the so-called *transuranic* elements—have been produced artificially in the laboratory. Although the search for a stable transuranic atom goes on, none have been discovered so far.

In microscopic terms, the difference in chemical behavior between any two distinct atoms has to do with the varying number of electrons which orbit about their respective atomic nuclei. The *atomic number* Z of an element is defined to be this number of electrons that orbit the nucleus. Equivalently, since the (positive) nuclear charge itself determines the number of electrons in orbit, the atomic number Z also represents this number of units of electric charge on the nucleus. For example, since the nucleus of the hydrogen atom has one unit of charge, it has one electron in orbit and its atomic number is unity. Similarly, helium has two electrons orbiting about it, and hence has an atomic number 2. Continuing on, we find lithium with an atomic number 3, beryllium with an atomic number 4, boron with atomic number 5, and so on, to the element uranium with atomic number 92.

A second important characterization of an atom is the *mass number, A.* As is the atomic number Z, the mass numbers of atoms are also positive integers. As the name implies, the mass number has to do with the mass of

the individual atom. To an accuracy better than 1 percent, the ratio of the mass numbers of two atoms is the same as the ratio of their actual masses. However, in discussing the atomic masses of naturally occurring elements, care must be exercised since for a given element—that is, one corresponding to a fixed value for the atomic numbers Z—a number of different species corresponding to different mass numbers A may exist. We call these various species *isotopes* of that element. For example, helium occurs in nature as a mixture of two isotopes; one with a mass number 3 and the other with a mass number 4. The isotopes of an element are distinguished by writing as a superscript the mass number A at the upper left of the symbol representing the chemical element. Thus the symbols ^3He and ^4He are used to represent the two stable isotopes of helium. Similarly, the atomic number of oxygen is 8, and it occurs naturally as a mixture of the three isotopes ^{16}O, ^{17}O, and ^{18}O, which have the mass numbers 16, 17, and 18, respectively.

The actual mass of an atom is usually expressed in terms of a unit known as the *atomic mass unit* or amu. The amu is defined so that the isotope of carbon ^{12}C, which is the lighter of the two isotopes ^{12}C and ^{13}C of naturally occurring carbon, has a mass of precisely 12 amu. Thus, by definition,

$$m(^{12}C) = 12 \text{ amu}$$

where $m(^{12}C)$ is the mass of the ^{12}C atom. With this definition, we find that to an accuracy of better than one percent the mass of the isotope of any element is numerically equal to its mass number A in atomic mass units.

A quantity related to the mass number of an element is its *atomic weight.* This is defined to be the mass of the atom of any element averaged over the various isotopes in which it is found to occur naturally. It is for this reason that not all atomic weights are integers in units of the amu. For example, the element chlorine is found to consist of a mixture of the two isotopes ^{35}Cl, and ^{37}Cl in the relative abundance of 3 to 1. Hence the atomic weight of chlorine is 0.75×35 amu $+ 0.25 \times 37$ amu $= 35.5$ amu. Similarly, we find for the lightest four elements, hydrogen, helium, lithium, and beryllium, the respective atomic weights of 1.01, 4.00, 6.94, and 9.01, all in units of amu.

Except for situations for which the existence of the isotopes of an element plays an important role, it is customary to omit writing the mass-number superscript next to the symbol for the chemical element. Thus we shall write "He" without the superscript when referring to helium in its naturally occurring state.

The relation between the amu and the macroscopic unit of the kilogram is of obvious physical interest and importance. To state this relation let us define the unit of the *mole* to represent the amount of a substance equal to its atomic weight in grams. Thus, since its atomic weight is 4.0, one mole of helium has a mass of 4.0 grams. Similarly, one mole of water (H_2O) has a mass of 18 grams, since the atomic weight of hydrogen is 1.0 and that of oxygen is nearly 16. The atomic hypothesis implies that the number of molecules in a mole must be the same for all substances; this number is

known as *Avogadro's number.* Experiment shows that it has the value

$$N_0 = 6.0222 \times 10^{23} \text{ atoms/mole} \tag{1-1}$$

By definition of a mole, the actual mass of an atom in grams is its atomic weight divided by Avogadro's number. For example, since one mole of helium has a mass of 4.0 grams, it follows that the mass of a helium atom $m(\text{He})$ has the value

$$m(\text{He}) = \frac{4.0 \text{ g/mole}}{6.0 \times 10^{23} \text{ atoms/mole}} = 6.7 \times 10^{-24} \text{ g/atom}$$

The Periodic Table of the elements appears at the back of the book following the Appendixes. It shows the familiar ordering of the elements in terms of their chemical properties and masses. Associated with each entry in the table is the chemical symbol for the element and two numbers. One of these numbers is an integer and stands directly above the chemical symbol; it represents the atomic number Z of the element. The second number is the atomic weight (in amu) of that element as it occurs naturally. In this grouping of the elements, which was originally proposed by Mendeleyef in 1870, elements with similar chemical properties appear in the same vertical column. For example, the entries in the first column are the alkaline metals, such as hydrogen, lithium, sodium, and potassium, whereas in column 7a we find the halogens: fluorine, chlorine, bromine, and so forth.

†1-11 Nuclear and subnuclear particles

Our present-day picture that an atom consists of a very small and positively charged nucleus surrounded by an appropriate number of electrons dates only from the year 1911, when Ernest Rutherford (1871–1937) first proposed it on experimental grounds. According to this *Rutherford model,* as it is known, almost all of the mass of the atom is concentrated in the nucleus, which has a diameter of the order of 10^{-15} meter. The size of the atom itself is about 100,000 times larger and is 10^{-10} meter in diameter. It is determined by the distance at which the outermost electrons circle the nucleus. The size of the atom, in other words, is determined by the distance from the nucleus of the outermost electrons in a manner analogous to the way the size of the solar system is fixed by the orbits of the outermost planets.

Figure 1-5 presents a sketch of the atoms of hydrogen and of helium according to this model. The hydrogen atom consists of a nucleus (having one unit of positive charge) and an electron (with its compensating negative charge) orbiting at a distance of approximately 10^{-10} meter from the nucleus. For the second element in the periodic table, helium, we find a similar situation, but this time the nucleus has two units of positive charge and thus two electrons orbiting about it. More generally, for an atom of atomic

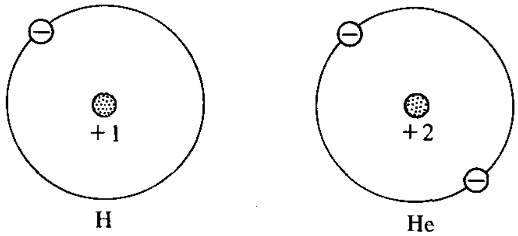


Figure 1-5

number Z the nucleus has Z units of positive charge and Z electrons orbiting about it.

With the concept of the indivisibility of atoms destroyed, the nucleus of the atom itself became the subject of close scrutiny. It was found that atomic nuclei are themselves not indivisible, but are composed of two basic entities: *neutrons* and *protons*, or, generically, *nucleons*. That is, it was found that the nucleus is composed of a number of positively charged protons and of uncharged neutrons and that the difference between two distinct nuclei and thus between the corresponding atoms is due entirely to the different number of nucleons of which they are composed. The proton itself is the nucleus of the hydrogen atom and has one unit of positive charge. The neutron is similar to the proton in most respects except that it is uncharged. The total number of nucleons in a nucleus defines the mass number A of the nucleus and therefore also of the corresponding atom. Although the neutron is slightly more massive than the proton, within an accuracy of 1 percent they have the same mass of 1.0 amu. Hence, to the same accuracy, the mass number A of the atom as measured in amu represents the mass of the atom. Note in this connection that since the mass of a proton is more than 1800 times that of an electron, to this accuracy the presence of electrons plays no role in determining the atomic mass.

Figure 1-6 shows schematically the composition of the various isotopes of the lightest two elements. The addition of a neutron to the nucleus of hydrogen changes its mass number A from 1 to 2 and in this way it becomes the nucleus of deuterium, or ^{2}H. Similarly, on adding a second neutron to this nucleus we obtain the radioactive (unstable) nucleus of tritium, or ^{3}H, with mass number 3. In the same way, ^{4}He has atomic number 2 and mass number 4, and hence its nucleus consists of two protons and two neutrons. The removal of one of these neutrons produces a nucleus of mass number 3 and with the same chemical properties as helium. It is the isotope ^{3}He.

More recent studies, mainly during the past two decades, have shown that the neutron, the proton, and the electron may also not be the ultimate building blocks of matter. Table 1-2 lists the particles discovered before and during this period and some of their properties. They are divided into two classes: *hadrons*, or strongly interacting particles, and *leptons*, or weakly interacting particles. Included among the leptons are the electron and the muon neutrino (ν_e and ν_μ), the electron (e^-), and the muon (μ^-), as well as

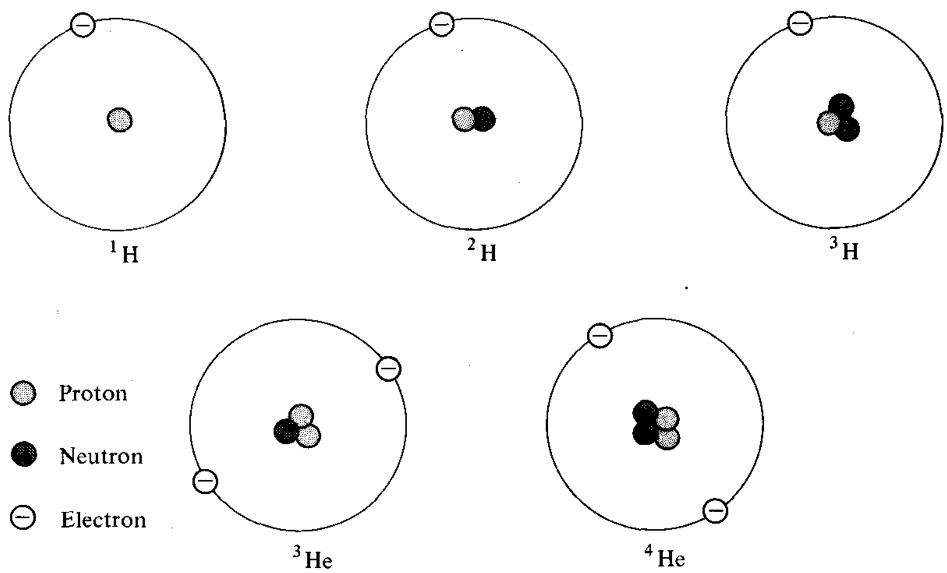


Figure 1-6

the corresponding antiparticles the antineutrinos ($\bar{\nu}_e$, $\bar{\nu}_\mu$), the *positron* (e^+), and the positive muon (μ^+). The *hadrons* themselves consist of two classes: the mesons and the baryons. As shown in the table, the particles classed as mesons are the three pions, π^+, π^-, π° (a positively charged one, its negatively charged antiparticle, and a neutral pion), and the kaons (consisting of a positively charged member K^+, a neutral member K°, and the corresponding antiparticles K^-, $\overline{K^\circ}$). The short-lived eta meson (η°) is also a member of this set. Nucleons belong to the second class of hadrons, and are known as *baryons*. Besides the nucleons, other members of the set are the lambda (Λ°); the sigma (Σ), which comes in three charge states ($\Sigma^+, \Sigma^-, \Sigma^\circ$); the negative and the neutral Xi (Ξ^-, Ξ°); and finally the omega minus (Ω^-). Each of these particles, as shown in the table, has its antiparticle of precisely the same mass but with opposite charge. With the exception of the electron, the proton, the neutrino, and the associated antiparticles, all of the remaining particles in this list are *unstable*. As a measure of this instability, the table lists the *lifetimes* of each of these unstable particles in the column labeled *mean life*. The mean-life is a measure of how long a particle lives prior to its decay. Note that the neutron is itself unstable and decays with a mean-life of approximately 15 min. This particular decay refers, of course, to neutrons in the free state and not to a neutron when it is bound inside the nucleus. Also listed in the table are certain decay modes associated with the unstable particles. For example, the decay mode of the neutron is

$$n \rightarrow p + e^- + \bar{\nu}_e$$

Similarly in a time interval of the order of 10^{-10} second, the Ω^- decays as follows:

$$\Omega^- \rightarrow \Lambda^\circ + K^-$$

Table 1-2

Particle	Symbol	Mass (units of electron mass)	Charge (proton charge units)	Mean life (s)	Decay products (example)	Antiparticle
Leptons						
Neutrino	ν_e, ν_μ	0	0	Stable	—	$\bar{\nu}_e, \bar{\nu}_\mu$
Electron	e^-	1	−1	Stable	—	e^+ (positron)
Muon	μ^-	207	−1	2.2×10^{-6}	$(e^-, \nu_\mu \bar{\nu}_e)$	μ^+
Hadrons						
Mesons						
Pion	π^+	274	+1	2.6×10^{-8}	(μ^+, ν_μ)	π^-
	π°	264	0	8.4×10^{-17}	(2γ)	π°
Kaon	K^+	967	+1	1.2×10^{-8}	(μ^+, ν_μ)	K^-
	K°	975	0	0.88×10^{-10}	(π^+, π^-)	$\bar{K}^\circ$
η-meson	η°	1074	0	$\sim 10^{-18}$	(2γ)	η°
Baryons						
Nucleons						
Proton	p	1836	+1	Stable	—	$\bar{p}$
Neutron	n	1839	0	918	$(p, e^-, \bar{\nu}_e)$	$\bar{n}$
Hyperons						
Lambda	Λ°	2184	0	2.5×10^{-10}	(p, π^-)	$\bar{\Lambda}^\circ$
Sigma	Σ^+	2327	+1	8.0×10^{-11}	(p, π°)	$\bar{\Sigma}^+$
	Σ°	2333	0	$< 1.0 \times 10^{-14}$	(Λ°, γ)	$\bar{\Sigma}^\circ$
	Σ^-	2342	−1	1.5×10^{-10}	(n, π^-)	$\bar{\Sigma}^-$
Xi	Ξ°	2573	0	3.0×10^{-10}	$(\Lambda^\circ, \pi^\circ)$	$\bar{\Xi}^\circ$
	Ξ^-	2585	−1	1.7×10^{-10}	(Λ°, π^-)	$\bar{\Xi}^-$
Omega	Ω^-	3275	−1	1.3×10^{-10}	(Λ°, K^-)	$\bar{\Omega}^-$

Figure 1-7 is a hydrogen bubble-chamber photograph taken at the Brookhaven National Laboratory in which a very fast π^- collides with a proton to produce a positive and a neutral kaon and a π^- and a neutron (K^+, $\bar{K}^\circ$, π^-, and n). Since only charged particles leave tracks in such a chamber the $\bar{K}^\circ$ and the n are invisible. Nevertheless, as shown by the dotted line in the diagram, we infer the existence of the $\bar{K}^\circ$ by virtue of its subsequent decay into a π^+-π^- pair. Similarly, the K^+ decays very rapidly into a π^+-π° pair. The existence of the neutral π° is inferred by the subsequent appearance of an electron positron (e^+-e^-) pair, which in turn is produced by one of the two gamma rays into which the π° decays. Note in the diagram that the

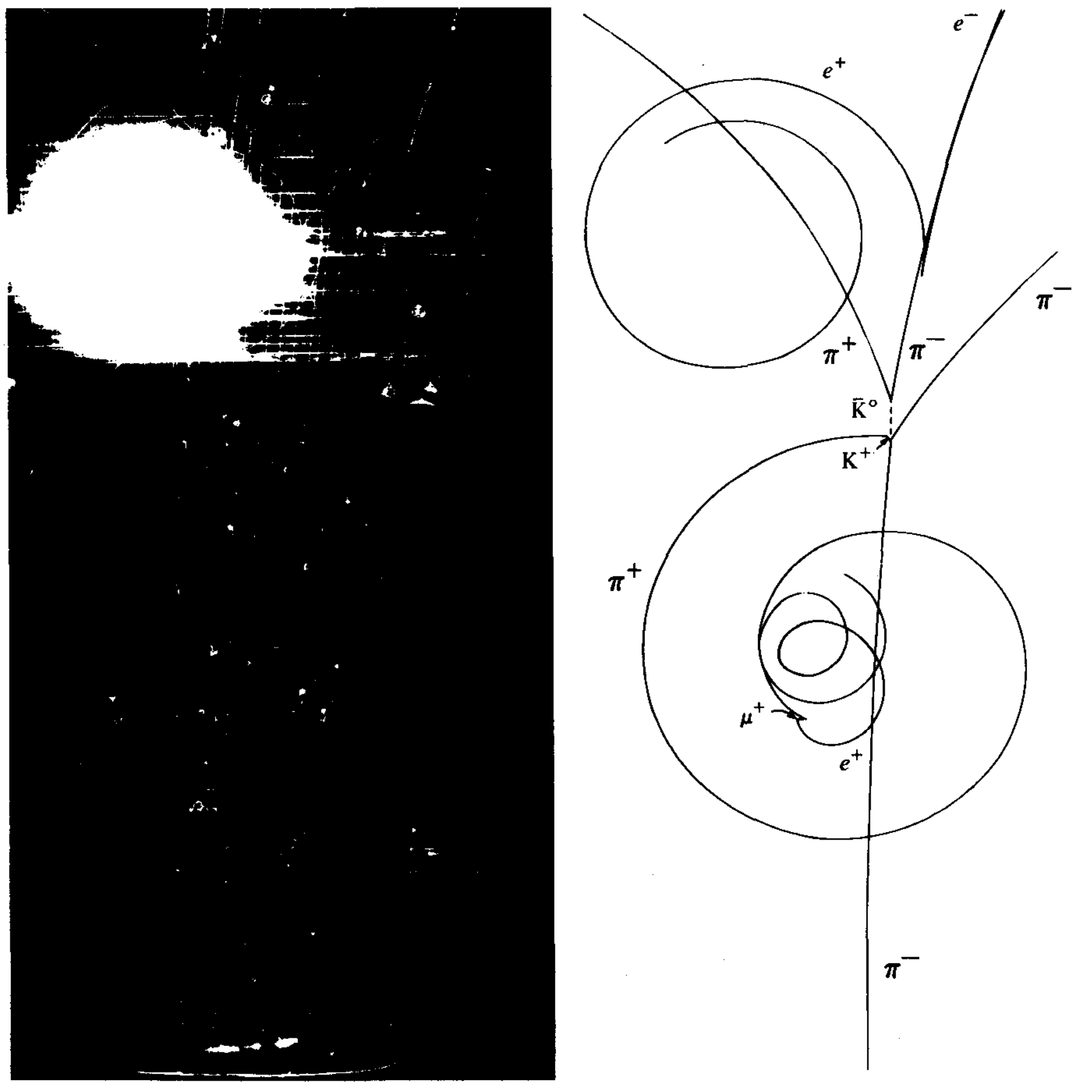
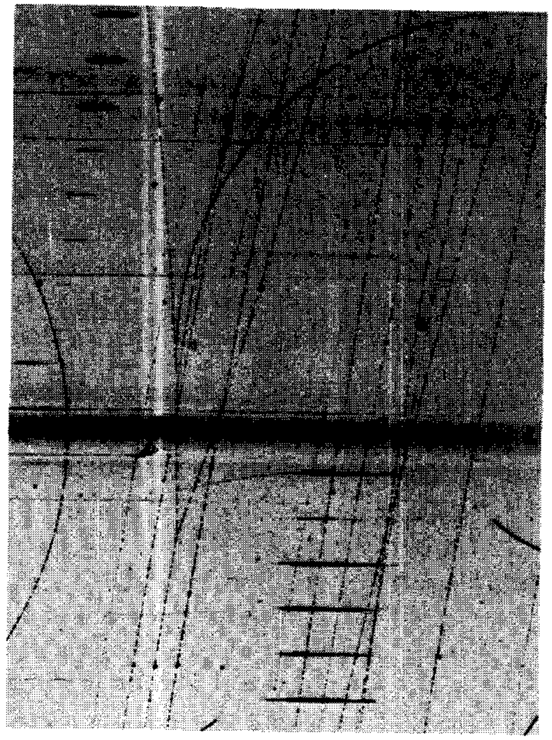


Figure 1-7 A hydrogen bubble-chamber photograph with an explanatory sketch showing the collision of a negative pion and a proton to produce two kaons, a neutron, and a negatively charged pion. Also shown are the subsequent decays of the neutral kaon into a π^+-π^- pair, and the K^+ into a π^+ and a π°. The existence of the neutral pion is inferred by the appearance of an electron-positron pair at the top of the diagram. (Courtesy Brookhaven National Laboratory.)



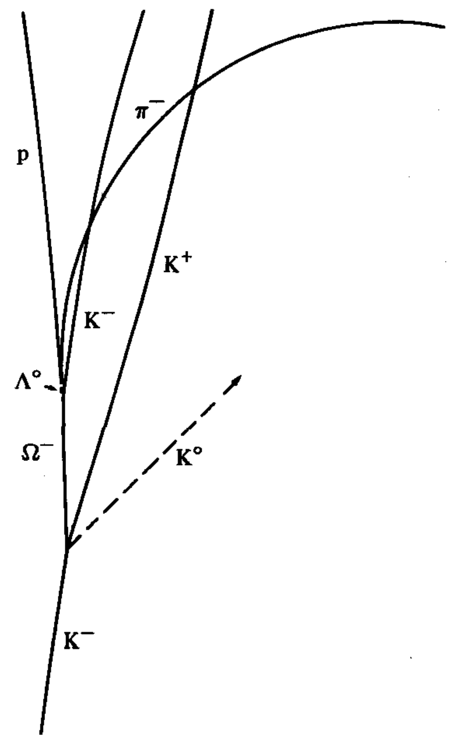


Figure 1-8 The hydrogen bubble-chamber photograph and an explanatory sketch which shows the third observation of the production of the Ω^-. The incoming K^- collides with a proton to produce a K^+, a K°, and an Ω^-. Also shown is the subsequent decay of the Ω^- into a K^- and a Λ° and the ensuing decay of the Λ° into a π^- and a proton. (Courtesy Brookhaven National Laboratory.)

tracks of all positively charged particles are bent to the left and those for negatively charged ones are bent to the right. This effect is produced by having a large magnet near the bubble chamber, which enables us to identify the charges of the particles producing the tracks.

Similarly, Figure 1-8 shows an observed event involving the production of the Ω^- at the Brookhaven National Laboratory. The reaction is

$$K^- + p \rightarrow \Omega^- + K^+ + K^\circ$$

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** units; **(b)** standard; **(c)** mole; and **(d)** atomic weight.
2. Which is longer, 100 yd or 100 meters? Is a mile longer or shorter than a kilometer, and by what factor?
3. What is the outstanding advantage that the metric system has over the English system? List also a secondary advantage.
4. The diameter of our sun is approximately 1.4×10^6 km. Give an operational meaning to this statement by finding out how this length is actually determined.
5. Suppose you have a meter stick with markings separated by 1 mm. Determine by its use the thickness of a

page of this book, and give an operational meaning to this length by describing your procedure.

6. The radius of a hydrogen atom is 5.3×10^{-11} meter. Give an operational meaning to this statement by assuming that hydrogen atoms are hard spheres of this radius, and thus predict the density—that is, the mass per unit volume—of hydrogen in the liquid form. Compare your answer with the density of liquid hydrogen in a handbook.
7. Explain why it was necessary, prior to 1960, to specify the temperature in defining the standard of length in terms of the distance between two lines on a certain platinum-iridium bar kept at the BIPM. Why is it not necessary to do the same thing for the prototype for the kilogram, which is also a platinum-iridium cylinder?
8. Give an operational meaning to the height of a very large tree by describing how this quantity can be measured by actually measuring the length of the shadow it casts.
9. Galileo, in carrying out some of his experiments in mechanics, used his own heartbeat as a way for measuring time. What are two disadvantages of this method? Are there any circumstances in which this might be advantageous?
10. What is the definition of Avogadro's number? Give an operational meaning to its experimental value of 6.02×10^{23} atoms/mole.
11. What is meant by the term "isotope"? Explain the deviations from integers of the atomic weights of some elements, such as chlorine, by assuming the existence of isotopes.
12. The "year" is the period of time associated with the earth's making a complete circuit about the sun; the "day" is the period of time required for the earth to make one full rotation about its axis; and the "month" is related to the time required for the moon to complete one orbit about the earth. Can you think of any relation of our seven-day week to the motion of some celestial body or bodies? (*Hint*: Think of the *names* of the days of the week.)

PROBLEMS

1. How many seconds are there in a year?
2. Calculate your height in meters.
3. You are driving through a country that uses the metric system and see a sign advising you that the speed limit is 90 km/hr. What is the corresponding value for this speed limit in the United States?
4. Using the facts that the mass of the sun is approximately 2.0×10^{30} kg and that its diameter is 1.39×10^{6} km, calculate its mean density—that is, its mass per unit volume.
5. Assuming that the earth is a sphere of radius 6.4×10^{3} km and has a mean density of 5.5×10^{3} kg/m^3, calculate the mass of the earth.
6. Experiments over the past decade have shown that the proton can be pictured as a sphere of radius 1.5×10^{-15} meter. Using the fact that the mass of the proton is 1.67×10^{-27} kg, calculate the density of the proton and compare your answer with the density of ordinary water.
7. Assuming that the length of the day increases by 10^{-3} second per century, by how much will the day 10^{6} yr from now be longer than the day today? Will this affect the length of the year?

8. The density of the element mercury at room temperature is $1.36 \times 10^4\ \mathrm{kg/m^3}$. **(a)** How many mercury atoms are there in a volume of $1\ \mathrm{m^3}$? **(b)** How many atoms are there in 1 gram of mercury?
9. How many atoms are there in a 1-kg platinum cylinder? Assume that the atomic weight of platinum is 195.
10. Assuming that the average atomic weight of the earth is 30 and using the data for the earth in Problem 5, calculate approximately the number of atoms in the earth.
11. If the average atomic weight of the air in the room in which you are now is 15 and its density is $1.29\ \mathrm{kg/m^3}$, what is the number of air molecules in the room?
12. Estimate the number of atoms in your body. Assume that the average atomic weight of the atoms in your body is 10.
13. Assuming that the water molecules in $1\ \mathrm{cm^3}$ of water are closely packed and that the density of water is $10^3\ \mathrm{kg/m^3}$, estimate the size of a water molecule.
14. Natural bromine exists in the two isotopic states of ^{79}Br and ^{81}Br, which have the respective atomic masses of 78.9 amu and 80.9 amu. If the atomic weight of bromine as given in the periodic table is 79.9, calculate the percentages of ^{79}Br and ^{81}Br in naturally occurring bromine.

†15. A neutron decays into a proton, an electron, and a neutrino. **(a)** Calculate the mass in grams lost in this decay process. Use the masses given in Table 1-2. **(b)** What has happened to this mass? Explain.

†16. Repeat Problem 15, but this time for the decay

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

of a positively charged pion (π^+) decaying into a similarly charged muon (μ^+) and a neutrino (ν_μ).

†17. **(a)** By use of the masses given in Table 1-2, verify the theoretically derived *Gell-Mann Okubo mass formula* for baryons:

$$\frac{1}{2}(M_\Sigma + 3M_\Lambda) = M_n + M_\Xi$$

where M_Σ is the average mass of the triplet Σ^+, Σ^-, Σ°; M_Λ is the mass of the lambda, and so forth.

(b) Calculate the mass of the eta particle m_η as given by the Gell-Mann Okubo formula for mesons:

$$\frac{1}{4}(m_\pi^2 + 3m_\eta^2) = m_K^2$$

where m_K is the mass of the kaon, and so forth. Compare with the value in Table 1-2.

†18. Confirm by reference to Table 1-2 that the ratio of the proton to the electron mass is precisely $6\pi^5$.

2 Kinematics in one dimension

The essential fact is that all the pictures which science now draws of nature... are mathematical pictures.

SIR JAMES JEANS (1877–1946)

2-1 Introduction

With this chapter we begin our study of classical physics by considering first one of its oldest and most basic branches—the science of *mechanics.*

Mechanics is the study of the motion of material objects and of the influences or forces that produce changes in motion. The importance of this discipline may be judged by the fact that many physical phenomena—for example, the mutual collisions of the molecules in a gas, the speeding up of protons in a cyclotron, the docking in a lunar orbit of two spacecraft, and so forth—involve motion of some sort. According to some points of view, it is virtually impossible to understand many fields of science without first acquiring some knowledge of mechanics.

The subject matter of mechanics is conveniently broken down into two subdivisions: *kinematics* and *dynamics.* Kinematics is that branch of mechanics concerned only with the description of motion—that is, the description of motion without regard to the causes that can affect or alter this motion. The science of dynamics, by contrast, deals with the complementary problem of correlating the motion of a material system with its structure and with the forces that can influence this motion. Thus, the

description of the motion of an automobile traveling along a highway, or that of an airplane flying between two cities, or that of a space probe on its way to Mars all come under the mantle of kinematics. If, in addition, we are interested in the forces by means of which these vehicles—starting, say, from rest—achieve their respective states of motion, then the problem is one of dynamics. For example, the problem of describing the historic descent to the lunar surface of Neil Armstrong and Edwin Aldrin in their Apollo 11 lunar module (see Figure 2-1) is a problem in kinematics. On the other hand, given the desired trajectory of this descent, the specification of when and for how long and with what strength the astronauts must operate the thrusters is a problem of dynamics.

We begin our studies of mechanics in this chapter by considering the kinematical problem of describing motion along a straight line. The related problem of motion in two and three dimensions will be the subject of Chapter 3.

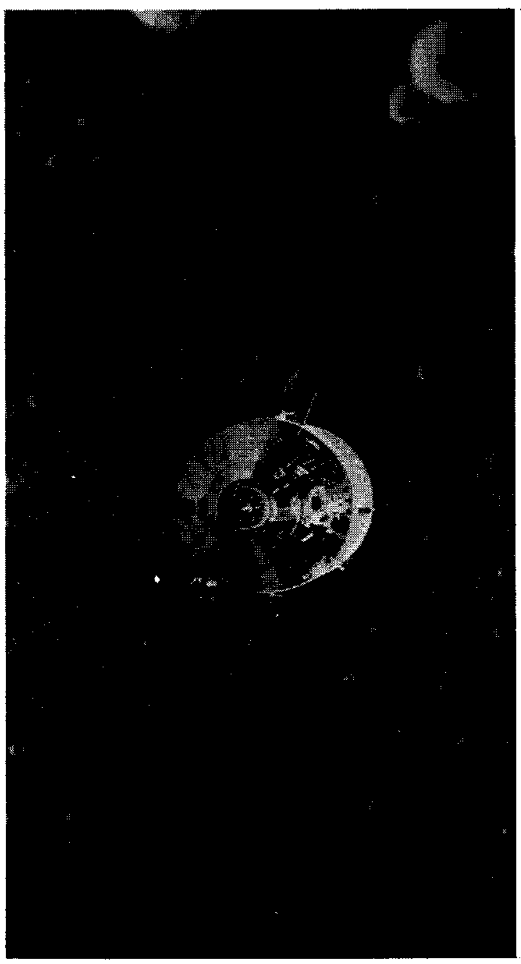


Figure 2-1 *The Apollo 11 Command Module with Michael Collins aboard as seen from the Lunar Module by Neil Armstrong and Edwin Aldrin as they prepare to descend to the lunar surface. The background shows the lunar terrain on the far side of the Moon. (Courtesy NASA.)*

2-2 Concept of a point particle

As a preliminary to a study of kinematics, it is appropriate to consider briefly the class of objects whose motion will be described. Although we shall in fact be concerned with the motions of a variety of systems ranging all the way from those of an individual atom or molecule to those of a planet or even a galaxy, it is very convenient initially to formulate the description of motion in terms of a certain idealized construct known as a point particle.

We define a *point particle*, or a *particle*, to be an object that has mass but no recognizable physical structure or extent. As far as we know today, such particles probably do not exist in nature. However, since the location of such a particle can be described very simply in terms of the coordinates of a geometric point, the notion itself is very useful and convenient. Furthermore, any real physical system, such as a large organic molecule or a planet, can be thought of as consisting of a very large number of these particles, so the motion of such a system can be described in terms of the individual motions of its constituent particles.

The main arguments justifying the usage of the notion of a point particle is that it works, it is very useful, and, most important, it can be employed to predict results consistent with experiment. Furthermore, as we shall see in Chapter 11, associated with *every* physical system is a certain point, called its *center of mass*, whose motion characterizes that of the system as a whole. In this sense the motion of any system may often be thought of in terms of the motion of this associated "point particle." For example, since the center of mass of the earth is at its center, the motion of the earth about the sun is usually described in terms of that of an "equivalent" point particle located at the earth's center. As a matter of fact, this is what is normally meant when we talk of the motion of any planet about the sun. To be sure, such a description is incomplete. For we know that the earth also rotates about its own axis, and thus in general at any instant no two points of the earth undergo precisely the same motion. Nevertheless, the description of the translational motion of the earth about the sun in terms of the motion of its center is often very convenient.

A second approach that may be used to justify the concept of a point particle is in terms of a *model*. According to this view, which is widely used in many areas of research, this idealized concept of a particle without extension is created in order to construct, by its usage, a detailed model that can be used in a practical way to make predictions in the physical world. Let us consider, in this connection, the concept of the triangle of Euclidean geometry. Brief reflection shows that no one has ever really seen a triangle, except possibly in his mind's eye. The triangles we draw on paper or see in books consist of straight line segments of finite thickness, while those of Euclid consist of lines of strictly zero thickness. Many theorems dealing with these Euclidean triangles have been developed, and we know by

experience that these theorems are also applicable, to a reasonable degree of accuracy, to the actual physical triangles drawn on paper. In this sense the triangles of Euclid serve as a model for the crude pseudotriangles actually seen in nature. In a similar sense, the point particles used in mechanics can be thought of as providing us with a model which can be used to describe actual "particles," such as atoms, molecules, planets, galaxies, and so forth, that do exist in nature.

2-3 Coordinate systems

By the description of the motion of a particle we shall mean a statement or a series of statements in terms of which the location of the particle is specified at any time t. In this connection it is convenient to introduce at this point the notion of a *coordinate system.* A coordinate system, it should be emphasized at the outset, is a purely mathematical construct that is useful for describing the position in space of a particle; it has no physical reality in itself.

Figure 2-2 shows a three-dimensional Cartesian coordinate system, consisting of the three mutually perpendicular axes: the x-, the y-, and the z-axes. If the point P represents the location of a particle, the position in space of this particle may be described uniquely in terms of the three directed, perpendicular distances x, y, and z from the various coordinate planes to the point P. As shown, the distances of the particle from the y-z, the x-z, and x-y planes are given respectively by the distances x, y, and z, which define the *coordinates* of the point P. We shall use the symbol (x, y, z) to represent the coordinates of this point.

In this chapter we are concerned only with motion in one dimension. For this case the particle is presumed to travel along a straight line, so only one coordinate is necessary to specify its position. Let us arbitrarily select the x-axis to be the direction along which the particle moves, so the location of the particle is given uniquely by the single coordinate x. As shown in Figure 2-3, for this case the particle travels in a direction parallel to the x-axis; thus the specification of the x-coordinate of the particle at any time t constitutes a complete description of its one-dimensional motion.

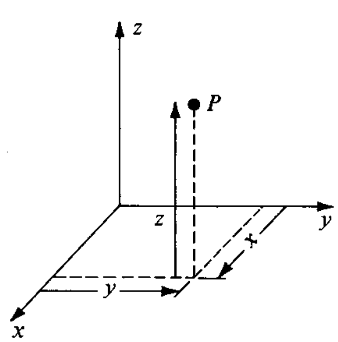


Figure 2-2

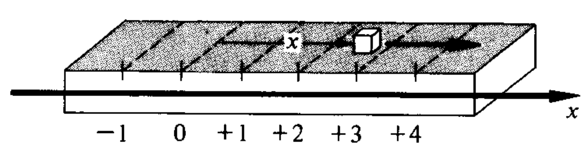


Figure 2-3

In selecting a coordinate system we have available a considerable amount of freedom with regard to both the algebraic sense of the axis and the choice of origin. Only after both of these quantities have been specified has the system been completely defined. As a rule, both the choice of origin and the algebraic sense of the axis are arbitrary. In Figure 2-3, for example, the origin of the x-axis is labeled by the symbol 0 and distances to the left of the origin have been arbitrarily selected to be negative. The opposite choice with distance to the left positive is equally acceptable. Also the distances marked off along the axis must be in units of length, such as meters or kilometers, and again, only after this unit of length along the axis has been specified has the coordinate system been completely defined.

What about the choice for the origin? Generally speaking it can be selected in any convenient way; there are no constraints on this choice whatsoever. If for example, it is desired to describe the vertical motion of an object that has been dropped from the top of a building, the origin can be taken at the top of the building, at the bottom of the building, or even at the center of the earth. Any such choice is acceptable, provided only that it is uniquely specified in advance and that the coordinate axis itself is selected to lie along the direction of motion. The complete specification of the coordinate system thus requires a choice for (1) the units in terms of which distances along the axis are measured; (2) its algebraic sense; and (3) the origin. Only after all arbitrariness has been removed in this way do we have available a coordinate system in terms of which the motion can be described.

2-4 Average velocity

Consider a particle moving in a direction parallel to the x-axis of a certain coordinate system. Because of its motion, the x-coordinate of the particle will vary in time; so we may write

$$x = x(t) \tag{2-1}$$

with $x(t)$ some function of time. For a given motion there is, as a general rule, a unique form for $x(t)$, and conversely.

Figure 2-4 shows a typical variation in time for such a *position function*

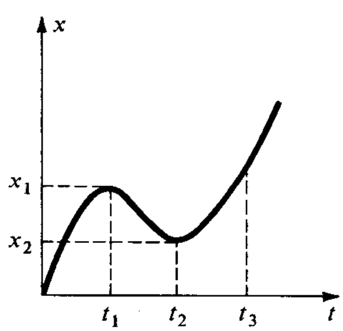


Figure 2-4

$x(t)$. The two axes in this plot, labeled x and t, are not to be confused with the underlying x-axis along which the particle is traveling. Note from the graph that the origin of the coordinate system has been selected to be the initial location of the particle at $t = 0$, and that the sense of the axis has been chosen to be positive along the initial direction of motion. For times $t < t_1$ the particle is traveling along the positive x-axis in the direction of increasing values for x, whereas for times t such that $t_1 < t < t_2$ it travels along the direction of decreasing values for x. Finally, for $t > t_2$ it resumes its original direction of motion. At the times t_1 and t_2, when the particle is located at the points x_1 and x_2, respectively, it is stationary, since at these points it is in the process of reversing its direction of motion.

The *average velocity* $\bar{v}$ of a particle undergoing motion is defined to be its net displacement in unit time. Specifically, if the particle travels along the x-axis and undergoes a displacement Δx in a time interval Δt, then its average velocity $\bar{v}$ is

$$\bar{v} = \frac{\Delta x}{\Delta t} \tag{2-2}$$

If $x(t)$ is the position of the particle at time t, then its position at the later instant $(t + \Delta t)$ will be $x(t + \Delta t)$ and its displacement Δx during this interval is therefore $[x(t + \Delta t) - x(t)]$. Hence an equivalent form for $\bar{v}$ is

$$\bar{v} = \frac{x(t + \Delta t) - x(t)}{\Delta t} \tag{2-3}$$

Note from this formula that the average velocity $\bar{v}$ can be both positive and negative. Thus if the particle travels along the positive sense of the x-axis, so that $x(t + \Delta t) > x(t)$, then (2-3) implies that $\bar{v} > 0$; conversely, if the particle moves along the direction of decreasing values for x, then $\bar{v} < 0$.

The meaning of (2-3) is further illustrated in Figure 2-5. The solid curve represents a part of a plot of the position function $x(t)$ of a particle. Reference to the definition in (2-3) and to the figure shows that the average velocity $\bar{v}$ is the same as is the tangent of the angle α in the figure. In particular, therefore, as α increases so does $\bar{v}$, and conversely. For the situation shown, if for example we decrease Δt, then both α and $\bar{v}$ must increase. Similarly, if Δt is fixed then as t varies, both the angle α and $\bar{v}$ will change. It follows that the

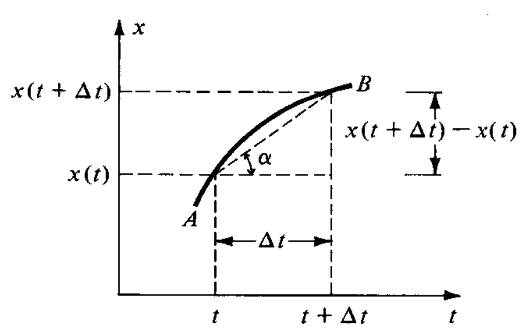


Figure 2-5

average velocity $\bar{v}$ of a particle, as defined in (2-2) or (2-3), will in general depend not only on time t, but on the time interval Δt as well.

Since the units for Δx are those of length, and the units of t and Δt are those for time, it follows from (2-2) that the units for $\bar{v}$, and for velocity in general, are those of length per unit time. In SI units, then, the unit of velocity is the meter per second, which we shall abbreviate as m/s.

Example 2-1 An automobile travels for a time interval of 1 min between two points a distance of 3 km apart. What is the average velocity?

Solution Substituting the given data, $\Delta x = 3.0 \text{ km} = 3.0 \times 10^3$ meters; $\Delta t = 1.0 \text{ min} = 60$ seconds into (2-2), we find that

$$\bar{v} = \frac{\Delta x}{\Delta t} = \frac{3.0 \times 10^3 \text{ m}}{60 \text{ s}} = 50 \text{ m/s}$$

Example 2-2 How long does it take an electron, which is traveling almost at the speed of light (3.0×10^8 m/s), to travel a distance of 1 mi along a linear accelerator?

Solution This time, the parameter values are

$$\bar{v} = 3.0 \times 10^8 \text{ m/s}$$

$$\Delta x = 1 \text{ mi} = 1.0 \text{ mi} \times \frac{1760 \text{ yd}}{\text{mi}} \times \frac{0.91 \text{ m}}{\text{yd}}$$

$$= 1.6 \times 10^3 \text{ m}$$

Solving (2-2) for Δt and substituting these values, we obtain

$$\Delta t = \frac{\Delta x}{\bar{v}} = \frac{1.6 \times 10^3 \text{ m}}{3.0 \times 10^8 \text{ m/s}} = 5.3 \times 10^{-6} \text{ s}$$

Example 2-3 The position $x(t)$ of a particle varies in accordance with the formula

$$x(t) = v_0 t$$

where v_0 is a positive constant. Calculate its average velocity $\bar{v}$.

Solution Substituting the given form for $x(t)$ into (2-3), we find that

$$\bar{v} = \frac{x(t + \Delta t) - x(t)}{\Delta t} = \frac{v_0(t + \Delta t) - v_0 t}{\Delta t} = v_0$$

Thus the average velocity $\bar{v}$ of the particle is independent of both t and Δt in this case. This type of motion, for which $x(t)$ depends linearly on t, is called *uniform motion.*

2-5 Instantaneous velocity

Except for the case of uniform motion, as in Example 2-3 for which $x(t)$ is a linear function of t, in general the average velocity $\bar{v}$ of a particle depends on the duration of the time interval Δt. Suppose, for example, that the position

$x(t)$ of a particle has the quadratic dependency

$$x(t) = 10t^2$$

where $x(t)$ is in meters and t is measured in seconds. According to (2-3), $\bar{v}$ is in this case given by

$$\begin{aligned}\bar{v} &= \frac{x(t+\Delta t) - x(t)}{\Delta t} = \frac{10(t+\Delta t)^2 - 10t^2}{\Delta t} \\ &= 20t + 10\Delta t \qquad \textbf{(2-4)}\end{aligned}$$

and thus depends on both t and Δt. Table 2-1 lists the values for $\bar{v}$ at the particular instant $t = 1$ second for various choices of Δt. Because of this kind of dependency of $\bar{v}$ on Δt, the notion of average velocity is not always a very convenient way to characterize the motion of a particle. A much more convenient and unambiguous one is its *instantaneous velocity.*

Table 2-1

Δt (seconds)	1	0.1	0.01	0.001	0.0001
$\bar{v}$ (meters per second)	30	21	20.1	20.01	20.001

The *instantaneous velocity*, or simply the *velocity* v, of a particle, is defined to be the limit of its average velocity at any instant in the limit as the time interval Δt (and its associated displacement Δx) tend to zero. Making use of the form for $\bar{v}$ in (2-3), we may express the instantaneous velocity $v = v(t)$ in the form

$$v(t) = \lim_{\Delta t \to 0} \bar{v} = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t} \qquad \textbf{(2-5)}$$

But the limit of the quotient here is, by definition, the derivative with respect to time, dx/dt, of the position function $x(t)$. Hence follows the alternate form

$$v(t) = \frac{dx}{dt} \qquad \textbf{(2-6)}$$

Thus the velocity of a particle may be obtained from its position function $x(t)$ simply by differentiation. Although it is possible, in principle, to calculate the velocity v of a particle directly by use of (2-5), usually (2-6) is the more convenient formula to use. A number of formulas useful for calculating derivatives are given in Appendix A.

Consider, for example, the situation described in (2-4). An application of (2-5) leads to the instantaneous velocity

$$v(t) = \lim_{\Delta t \to 0} (20t + 10\,\Delta t) = 20t$$

since as Δt tends to zero the term $10\,\Delta t$ vanishes, whereas the term $20t$ is fixed. In particular, at $t = 1$ second, we find that $v = 20$ m/s, which agrees with the limit of the sequence 30, 21, 20.1, 20.01, 20.001, in Table 2-1. Or equally, since for any constant α the time derivative of αt^2 is $2\alpha t$, the same result follows by applying (A-1), from Appendix A, directly to the position function $x(t) = 10t^2$.

In more physical language, we may describe the meaning of the instantaneous velocity of a particle in the following way. Suppose, in Figure 2-6, that at time t a particle is at the point x_0 and that at a later time $(t + \Delta t_1)$ it is at the point x_1. Then its average velocity $\bar{v}_1$ over this interval is $(x_1 - x_0)/\Delta t_1$. Let us now repeat this same experiment, but this time for a shorter time interval, Δt_2. In the notation of the figure, the average velocity $\bar{v}_2$ over this shorter interval is $(x_2 - x_0)/\Delta t_2$. Continuing on in this way we obtain a sequence of average velocities, $\bar{v}_1, \bar{v}_2, \bar{v}_3, \ldots$, corresponding to successively smaller time intervals and displacements. On physical grounds, this sequence must approach a limit, and it is this limiting value that has been defined to be the instantaneous velocity.

An alternate way of viewing this limiting process is shown in Figure 2-7. In the limit as the sequence of time intervals $\Delta t_1, \Delta t_2, \ldots$ tends to zero, the sequence of average velocities $\bar{v}_1, \bar{v}_2, \ldots$ approaches the same limit, as does the sequence of ratios

$$\frac{C_1B_1}{AC_1}, \frac{C_2B_2}{AC_2}, \frac{C_3B_3}{AC_3}, \ldots$$

where, for example, $C_1B_1 = x_1 - x_0$ is the distance traveled by the particle in the time interval $\Delta t_1 = AC_1$. Reference to the figure shows that this latter sequence approaches, in the limit, the slope of the $x(t)$ curve at point A. In other words: *the instantaneous velocity $v(t)$ of a particle at time t is equal to the slope of its position curve $x(t)$ at the time t.*

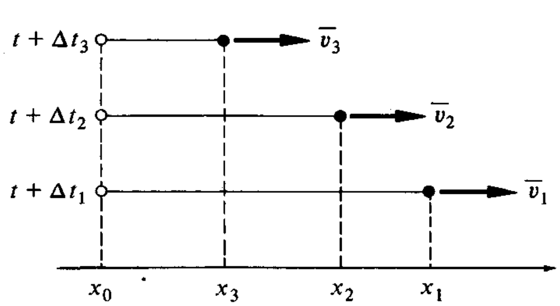


Figure 2-6

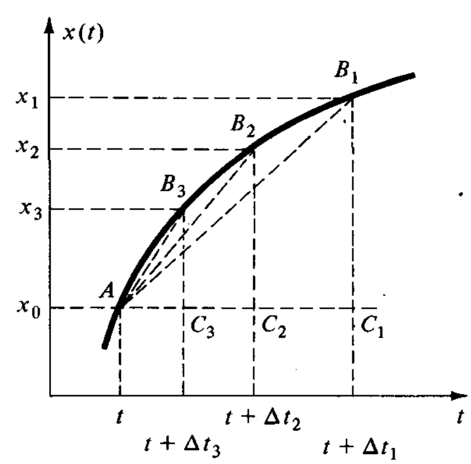


Figure 2-7

Thus the larger is the velocity of a particle, the greater is the corresponding slope on the $x(t)$ curve. For the special curve in Figure 2-4, for example, since for $t < t_1$ the slope is positive, it follows that the corresponding velocity of the particle is positive. At the points x_1 and x_2 the slope of the curve is zero, so the velocity of the particle vanishes at the times $t = t_1$ and $t = t_2$. Finally, for times t for which $t_1 < t < t_2$ the slope is negative, and this means that the velocity of the particle is also negative. In other words, in this region the particle travels in the direction of decreasing values for x.

Example 2-4 Calculate the velocity at any time t of a particle whose position $x(t)$ is given by

$$x(t) = x_0 + v_0 t + \frac{1}{2} a_0 t^2$$

where x_0, v_0, and a_0 are positive constants.

Solution Substitution into (2-3) yields the average velocity $\bar{v}(t)$ at time t:

$$\bar{v}(t) = \frac{x(t+\Delta t) - x(t)}{\Delta t} = \frac{x_0 + v_0(t+\Delta t) + \frac{1}{2}a_0(t+\Delta t)^2 - x_0 - v_0 t - \frac{1}{2}a_0 t^2}{\Delta t}$$

$$= v_0 + a_0 t + \frac{a_0 \Delta t}{2}$$

Taking the limit as $\Delta t \to 0$ we find the desired result

$$v(t) = \lim_{\Delta t \to 0} \bar{v}(t) = v_0 + at$$

For the special case, $v_0 = 0$, $a_0 = 20\ \text{m/s}^2$, this reduces to the formula considered earlier in this section.

Example 2-5 If the position $x(t)$ of a particle at time t is

$$x(t) = bt^3$$

with $b = 1.5\ \text{m/s}^3$, calculate the velocity of the particle at time $t = 2.0$ seconds.

Solution Making use of (2-5), we find

$$v(t) = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{b(t+\Delta t)^3 - bt^3}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{b[3t^2 \Delta t + 3t(\Delta t)^2 + (\Delta t)^3]}{\Delta t} = \lim_{\Delta t \to 0} b[(3t^2 + 3t\Delta t + (\Delta t)^2]$$

$$= 3bt^2$$

In particular, for $t = 2.0$ seconds this leads to

$$v(2.0\text{ s}) = 3 \times (1.5\text{ m/s}^3) \times (2.0\text{ s})^2 = 18\text{ m/s}$$

The same result also follows directly by use of (A-1), since the derivative of bt^3 is $3bt^2$.

2-6 Acceleration

If the velocity of a particle changes in any manner, then the particle is said to *accelerate* or to undergo an *acceleration.* An automobile that starts from rest, for example, and achieves a velocity of 90 km/hr, accelerates during this speeding-up process. Similarly, an electron accelerates when it collides with the electrode in an X-ray tube and thereby slows down. The purpose of this section is to describe this very important notion of acceleration in quantitative terms.

As in the analogous discussion of velocity, it is convenient first to define a quantity known as the *average acceleration.* Consider a particle, which at a fixed instant t is moving along a straight line at a velocity $v(t)$. A short time interval Δt later, its velocity will be $v(t + \Delta t)$. By analogy to (2-3), we define its average acceleration $\bar{a}(t)$ over the time interval Δt to be

$$\bar{a}(t) = \frac{v(t + \Delta t) - v(t)}{\Delta t} \tag{2-7}$$

In words, this definition states that the average acceleration of a particle is equal to the change in its velocity per unit time during a time interval Δt. In particular, if the velocity of the particle does not change in time, then it does not accelerate. We say in this case that it travels with *uniform velocity.*

As in the discussion of velocity, the average acceleration $\bar{a}$ as defined in (2-7) is manifestly dependent on the length of the time interval Δt and, in general, varies as Δt does. Accordingly, we define the *instantaneous acceleration* or simply the *acceleration,* $a(t)$, of a particle to be the limit of the average acceleration $\bar{a}(t)$ as Δt tends to zero; that is,

$$a(t) = \lim_{\Delta t \to 0} \bar{a}(t) = \lim_{\Delta t \to 0} \frac{v(t + \Delta t) - v(t)}{\Delta t} \tag{2-8}$$

where the second equality follows from (2-7). Since the limiting value of the quotient is, by definition, the derivative of $v(t)$, we have the equivalent form

$$a(t) = \frac{dv}{dt} \tag{2-9}$$

It follows that if the velocity $v(t)$ of the particle is known at any time t, then its acceleration may be computed directly by differentiation. If, instead, the position coordinate $x(t)$ of the particle is known, then according to (2-6) and (2-9) two differentiations are required: one to obtain $v(t)$ and the second to obtain $a(t)$.

It follows from either (2-8) or (2-9) that the units of acceleration are those of velocity per unit time. Thus, in SI units, the appropriate unit of acceleration is the meter per second squared (m/s^2).

It is important to note that regardless of the direction along which the particle travels, if its velocity increases, so that $v(t+\Delta t) > v(t)$, then, according to (2-8), its acceleration is positive. Similarly, when it slows down and $v(t+\Delta t) < v(t)$, its acceleration is negative. These features are illustrated in the graph of velocity versus time in Figure 2-8. Just as for the analogous discussion of $x(t)$ versus t in Section 2-5, the slope of the $v(t)$ curve is at each instant the acceleration of the particle. Thus, from the fact that for $t < t_1$ the slope is positive and that $v > 0$, it follows that the particle travels in the direction of increasing values for x and its velocity increases. Similarly, during the time interval $t_1 < t < t_2$ we have $v > 0$ and $a < 0$ since here the slope is negative, so as the particle travels in the direction of increasing values for x its velocity decreases. On the other hand, since $v < 0$ and $a < 0$ along the interval $t_2 < t < t_3$, it follows that the particle is moving in the direction of decreasing values for x with an increasing speed. At the times t_2 and t_4 the velocity of the particle vanishes; in other words, at $t = t_2$ and $t = t_4$ the curve of $x(t)$ versus t has a horizontal slope. In the same way, since at the times t_1 and t_3 the $v(t)$ curve has zero slope, we find that at these two times the acceleration of the particle vanishes.

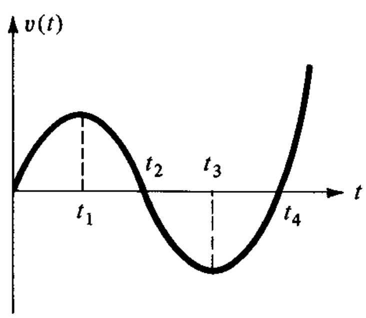


Figure 2-8

Example 2-6 Calculate the acceleration of the particle whose position coordinate is

$$x(t) = bt^3$$

with b a positive constant.

Solution According to the result of Example 2-5, the velocity $v(t)$ of the particle is

$$v(t) = 3bt^2$$

Substituting this form into (2-9) and making use of the differentiation formula in (A-1), we find that

$$a = \frac{dv}{dt} = \frac{d}{dt}(3bt^2) = 6bt$$

Alternatively, making use of (2-8), we obtain

$$a = \lim_{\Delta t \to 0} \frac{3b(t+\Delta t)^2 - 3bt^2}{\Delta t} = \lim_{\Delta t \to 0} \frac{3b[2t\Delta t + (\Delta t)^2]}{\Delta t}$$
$$= 6bt$$

Example 2-7 Show that the acceleration of the particle in Example 2-4, with the position function

$$x(t) = x_0 + v_0 t + \frac{1}{2} a_0 t^2$$

is independent of the parameters x_0 and v_0.

Solution Since x_0, v_0 and a_0 are constants, by substituting this form for $x(t)$ into (2-6) and making use of (A-1) we obtain

$$v(t) = \frac{d}{dt}\left(x_0 + v_0 t + \frac{1}{2} a_0 t^2\right)$$
$$= v_0 + a_0 t$$

A second differentiation then yields the acceleration

$$a(t) = \frac{dv}{dt} = \frac{d}{dt}(v_0 + a_0 t)$$
$$= a_0$$

and this is independent of the parameters v_0 and x_0.

2-7 Motion under constant acceleration—integration

If the position function $x(t)$ of a particle is known, then, according to (2-6) and (2-9), its velocity and acceleration may be obtained by taking the appropriate derivatives of x and v, respectively. A related problem of considerable importance in mechanics is the inverse of this problem; it involves obtaining the position function $x(t)$ given only the velocity or the acceleration of the particle. This process of extracting $x(t)$ from $v(t)$ or $a(t)$, or of $v(t)$ from $a(t)$, is known as *integration.* The purpose of this section is to illustrate the process of integration by reference to the special but important case of a particle undergoing constant acceleration.

Consider a particle in motion parallel to the x-axis of a certain coordinate system under constant acceleration a_0. Just as for $x(t)$ and $v(t)$, positive values for a_0 signify that the particle accelerates along the positive sense of the axis, and negative values along the opposite sense. We shall now establish that the position coordinate $x(t)$ and the velocity $v(t)$ of the particle are

$$x(t) = \frac{1}{2} a_0 t^2 + v_0 t + x_0 \tag{2-10}$$

$$v(t) = v_0 + a_0 t \tag{2-11}$$

with v_0 and x_0 completely *arbitrary* constants. If we set $t = 0$ in (2-10) and (2-11), it is apparent that x_0 represents the initial position of the particle and that v_0 is its initial velocity.

The validity of (2-10) and (2-11) follows directly from the definitions in (2-6) and (2-9) and the rules for differentiation in (A-1) and (A-2). First, since the time derivative of $(x_0 + v_0t + \frac{1}{2}a_0t^2)$ is $(v_0 + a_0t)$, it follows that the relations in (2-10) and (2-11) are consistent with (2-6). Moreover, since the derivative of $(v_0 + a_0t)$ is the constant acceleration a_0, these relations are consistent with (2-9) as well. Thus, regardless of the values for the arbitrary constants x_0 and v_0, (2-10) and (2-11) represent the position and velocity at any time t of a particle undergoing a constant acceleration a_0. These forms for $x(t)$ and $v(t)$ are also known as the *integrals* of the relation

$$\frac{dv}{dt} = a_0 \tag{2-12}$$

which, according to (2-9), is the formal statement that the acceleration of the particle is a_0.

It is important to note that even though for a given form for $x(t)$ there is in general a unique form for $v(t)$ and $a(t)$, the converse is not true. That is (as seen above for the case of constant acceleration), for a given $a(t)$ there are in general many position functions $x(t)$, and associated with each is its own velocity function. In other words, the process of differentiation is unique, whereas that of integration is not.

To illustrate the matter consider two particles, both of which have the same velocity $v(t)$ at all times t but are initially at a fixed separation distance b apart. It is evident physically, that under the stated conditions they will maintain this separation distance b for all times t. Hence if, as shown in Figure 2-9, the position coordinates of the two particles are $x_1(t)$ and $x_2(t)$, it follows that

$$x_2(t) - x_1(t) = b$$

But the velocities of the particles are precisely the same, and thus, according to (2-6),

$$\frac{dx_1}{dt} = \frac{dx_2}{dt}$$

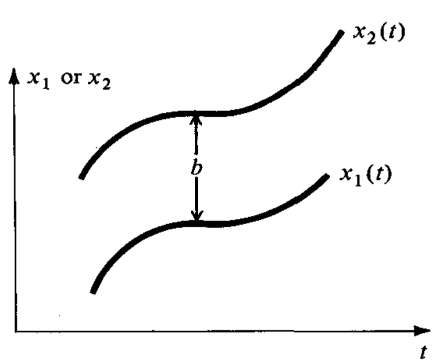


Figure 2-9

Hence the slopes of the two curves in Figure 2-9 have the same value at all times t. We have, therefore, a situation in which the velocities of two particles are everywhere the same but in which the associated position functions differ by an arbitrary constant. Hence follows the conclusion:

If the velocity $v(t)$ of a particle is specified for all t, then its position function is determined only up to an arbitrary additive constant.

As noted previously, the process of determining $x(t)$ from a given velocity $v(t)$, or of $v(t)$ from the acceleration $a(t)$, is known as *integration*. The arbitrary constant is called a *constant of integration*. Formally, we write the inverse of (2-6) as

$$x(t) = \int v(t)\, dt + c \tag{2-13}$$

which is to be read: $x(t)$ is equal to the (indefinite) integral of $v(t)$ plus an integration constant c. If we work with (2-9), the analogous relation is

$$v(t) = \int a(t)\, dt + c$$

It is important to bear in mind that integration and differentiation are *inverse processes*. To obtain the position function $x(t)$ for a given $v(t)$ we must ask ourselves the question: The derivative of what function yields the given form for $v(t)$? Once such a function is found, $x(t)$ is obtained by simply adding to it an arbitrary constant. If, for example, the known velocity has the form $v(t) = \alpha t^3$, then—since according to (A-1) the derivative of $\alpha t^4/4$ is αt^3—we have

$$x(t) = \int v(t)\, dt + c = \int \alpha t^3\, dt + c = \frac{\alpha t^4}{4} + c$$

Example 2-8 In a certain coordinate system the velocity of a particle is

$$v(t) = \beta t^3 - \alpha t^2$$

where β and α are two fixed constants. If at time $t = t_0$ the particle is located at the origin, determine the position of the particle at any time t.

Solution Since the integral of βt^3 is $\beta t^4/4$ and the integral of $-\alpha t^2$ is $-\alpha t^3/3$, it follows that $x(t)$ must have the form

$$x(t) = \frac{\beta}{4} t^4 - \frac{\alpha}{3} t^3 + c$$

where c is an arbitrary integration constant. Imposing the stated condition $x(t_0) = 0$, we find that

$$0 = \frac{\beta}{4} t_0^4 - \frac{\alpha}{3} t_0^3 + c$$

that is, the constant c is given by

$$c = \frac{\alpha}{3}t_0^3 - \frac{\beta}{4}t_0^4$$

Thus, the position of the particle is

$$x(t) = \frac{\beta}{4}(t^4 - t_0^4) - \frac{\alpha}{3}(t^3 - t_0^3)$$

Example 2-9 The acceleration of a particle increases linearly with time in accordance with the formula

$$a(t) = \alpha t$$

with α constant. If initially the particle is at the origin and has there a velocity v_0, find its velocity and position at any time t.

Solution Integrating the given relation

$$\frac{dv}{dt} = \alpha t$$

we find, since $d(\alpha t^2/2)/dt = \alpha t$, that

$$v(t) = \frac{\alpha}{2}t^2 + c_1$$

with c_1 a constant. Since $v = v_0$ at $t = 0$, it follows that $c_1 = v_0$ and thus

$$v(t) = \frac{\alpha}{2}t^2 + v_0$$

Integrating a second time we find, since the derivative of $\alpha t^3/6$ is $\alpha t^2/2$, that

$$x(t) = \frac{\alpha}{6}t^3 + v_0 t + c_2$$

Finally, since at $t = 0$ the particle is at the origin, so that $x = 0$, it follows that the constant c_2 vanishes, and thus

$$x(t) = \frac{\alpha}{6}t^3 + v_0 t$$

2-8 Free fall

An important application of the constant-acceleration formulas in (2-10) and (2-11) is to the case of a body falling near the surface of a planet such as the earth. As will be discussed in more detail in Chapter 4, in this case the body is under the influence of a uniform gravitational field and will move with a constant acceleration g, which is directed vertically downward and has the value

$$g = 9.8 \text{ m/s}^2 \qquad \textbf{(2-14)}$$

This corresponds in the English system to the value 32 ft/s^2.

Consider, in Figure 2-10, an object which is thrown vertically upward with an initial speed v_0 from a distance h above the ground. Let us set up a coordinate axis with the positive sense directed vertically upward and with the origin on the ground. In this system the initial position of the body has the coordinate h, and the body's acceleration is

$$\frac{dv}{dt} = -g \tag{2-15}$$

with g given in (2-14). The minus sign here reflects the fact that the acceleration of the body is directed vertically downward, so its sense is along the negative direction of the axis.

To obtain formulas for the position and velocity of the body at any time t, let us make, in (2-10) and (2-11), the substitutions $a_0 \to -g$ and $x_0 \to h$. The result is

$$x(t) = -\frac{1}{2}gt^2 + v_0 t + h \tag{2-16}$$

$$v(t) = v_0 - gt \tag{2-17}$$

These are the basic relations used to analyze motion in a uniform gravitational field. Eliminating the variable t between (2-16) and (2-17), we obtain the additional relation

$$v^2 = v_0^2 - 2g(x-h) \tag{2-18}$$

which is also of considerable practical interest.

Figure 2-11 shows a plot of (2-16). Since at $t = v_0/g$, the curve has zero slope, we see that the maximum height reached by the body is $h + v_0^2/2g$. From the symmetry of the curve it follows that the speed of the body when it is at a distance h above the ground during its downward trajectory is the same as its initial speed v_0. This also follows directly from (2-18), which shows that for $x = h$, $|v| = v_0$.

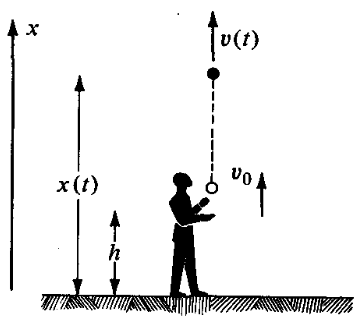


Figure 2-10

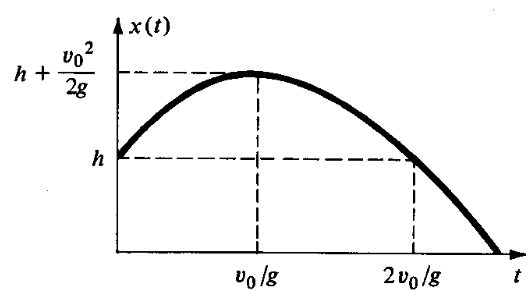


Figure 2-11

Example 2-10 Assume that you throw a stone vertically upward from the ground with an initial velocity of 20 m/s.

(a) How long does it take the stone to reach its maximum height?

(b) How high does it rise?

Solution The appropriate parameter values to be used in (2-16) through (2-18) in this case are $v_0 = 20$ m/s and $h = 0$.

(a) At the height of the trajectory the velocity v of the particle vanishes. Thus, if t_0 represents the instant when the particle is at its maximum height, then, according to (2-17),

$$0 = -gt_0 + v_0$$

or

$$t_0 = \frac{v_0}{g} = \frac{20 \text{ m/s}}{9.8 \text{ m/s}^2} = 2.0 \text{ s}$$

(b) Substituting this value for t_0 into (2-16), we obtain

$$\begin{aligned} x(t_0) &= -\frac{1}{2}gt_0^2 + v_0t_0 \\ &= -\frac{1}{2} \times (9.8 \text{ m/s}^2) \times (2.0 \text{ s})^2 + (20 \text{ m/s}) \times (2.0 \text{ s}) \\ &= 20 \text{ m} \end{aligned}$$

Example 2-11 A body is thrown vertically downward from a height of 30 meters above the ground with a certain downward velocity so that it strikes the ground 2.0 seconds after release.

(a) What was its initial velocity?

(b) With what velocity does it strike the ground?

Solution This time we are given the value $h = 30$ meters and asked to consider the situation at $t = 2.0$ seconds when $x = 0$.

(a) Substitution of these values into (2-16) yields

$$0 = -\frac{1}{2} \times (9.8 \text{ m/s}^2) \times (2.0 \text{ s})^2 + v_0 \times (2.0 \text{ s}) + 30 \text{ m}$$

Solving for v_0, we obtain

$$v_0 = -5.0 \text{ m/s}$$

The minus sign is a reflection of the fact that the initial velocity is downward and thus in a direction opposite to that of the coordinate axis in Figure 2-10.

(b) To find the final velocity of the body, let us make use of (2-18). Substituting the values $x = 0$ and $v_0 = -5$ m/s, we find that

$$\begin{aligned} v^2 &= v_0^2 - 2g(x-h) = (-5.0 \text{ m/s})^2 - 2 \times (9.8 \text{ m/s}^2) \times (0 - 30 \text{ m}) \\ &= 25 \text{ m}^2/\text{s}^2 + 590 \text{ m}^2/\text{s}^2 = 615 \text{ m}^2/\text{s}^2 \end{aligned}$$

Hence

$$v = -25 \text{ m/s}$$

where the choice of sign is for the same reason as above. This result can also be obtained by the substitution of $v_0 = -5.0$ m/s and $t = 2.0$ seconds directly into (2-17).

†2-9 Moving coordinate systems—the Galilean transformation

It is important to recognize the fact that the foregoing definitions for velocity and acceleration have involved in an essential way the use of a coordinate system. Since the velocity of a particle was defined in terms of the time derivative of its position function $x(t)$, and since $x(t)$ has meaning only with respect to a given system, it follows that implicit in these definitions is the assumption that a definite coordinate system has been established in advance. Moreover, in that discussion it was assumed implicitly that this coordinate system was invariably at rest relative to the observer for whom the description of motion was being carried out; that is, it was assumed that there was no relative motion between the observer and the origin of the coordinate system in terms of which the particle's position $x(t)$ had been defined. However, there do arise situations for which this latter assumption is not appropriate, and the purpose of this section is to discuss how the descriptions of the motion of a particle as seen by two observers in relative motion are related to each other.

As an illustration let us consider a train that is traveling due east with a velocity u, and suppose that a passenger on this train walks down the aisle, in the direction of motion of the train, at the velocity v_0. From the viewpoint of the passenger, it is natural to take a coordinate system fixed relative to the train, and in this system his motion is best described by saying that he travels down the aisle at the velocity v_0. Figure 2-12a shows this motion, which is observed by those who are at rest relative to the train. Consider now this same situation, but this time from the viewpoint of an external observer who is fixed relative to the earth, so that to him the train travels east at the velocity u. This external observer would naturally take a coordinate system at rest with respect to himself, and from this point of view the passenger appears to travel in the same direction as the train but at the velocity $(v_0 + u)$. That is, the observed velocity of the passenger is given this time by the algebraic sum of his velocity relative to the train, and the velocity of the train relative to the external observer. Figure 2-12b shows the situation as seen by this fixed observer. Note that these two descriptions—in one of

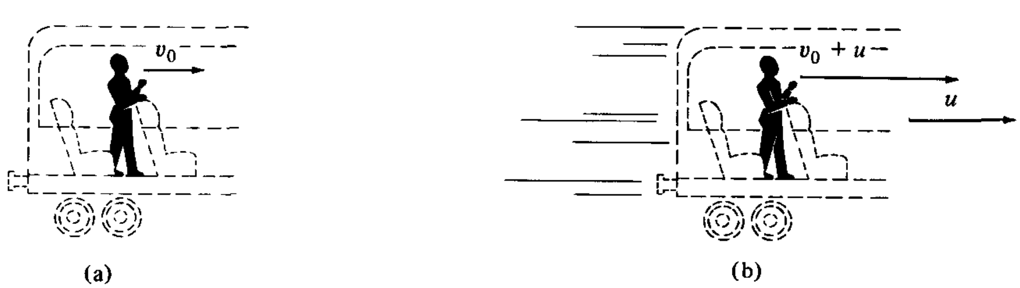


Figure 2-12

which the passenger is observed to travel east at the velocity v_0, while in the other he is observed to travel at the velocity $(u + v_0)$—correspond to *precisely the same physical situation,* but as described in two different languages. That is, each observer in describing this motion makes use of the language appropriate to his own choice of coordinate axis. This example illustrates clearly that in order to specify uniquely a coordinate system, it is necessary to specify not only its origin and the algebraic sense of the axis, but also to state with respect to what observer the system is stationary. Only in this way does a uniquely defined description of the motion of a particle result.

Consider, in Figure 2-13, two coordinate systems S and S', and suppose that S' travels relative to S with the constant velocity u along their parallel x- and x'-axes. This means that from the viewpoint of an observer fixed in S, the origin of S' appears to travel at the constant velocity u along the x-axis. Equivalently, from the viewpoint of an observer at rest in S' the origin of S travels at the constant velocity $-u$ along the x'-axis. For diagrammatic convenience only, the y- and y'-axes have been included and drawn perpendicular to the x- and x'-axes at their respective origins.

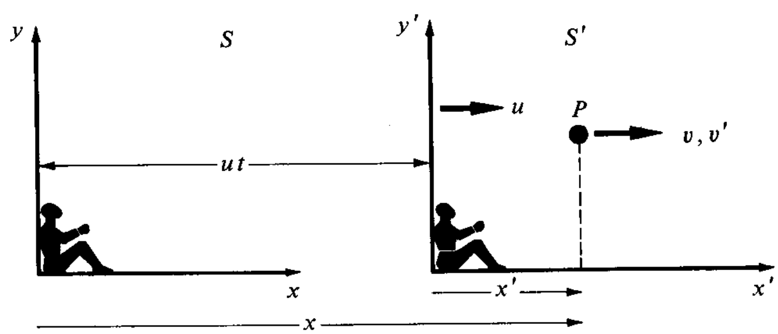


Figure 2-13

Suppose now that point P in the figure represents the instantaneous position of a particle that is traveling along the parallel x- and x'-axes. The observer in S' describes the instantaneous position of the particle at time t by the coordinate x', that is, by its perpendicular distance from the y'-axis, while the observer in S describes its position at this same instant in terms of its distance x from the y-axis. Assuming that the origins of the two systems coincide at $t = 0$, it follows that at time t the separation distance between them is ut. Hence reference to Figure 2-13 shows that these two equivalent ways for describing the position of the particle must be related by

$$x' = x - ut \tag{2-19}$$

We emphasize that the coordinates x' and x in this relation both refer to the same physical *point P*—namely, that point at which the particle is located at the instant t.

This relation in (2-19), which relates the two different ways for describing the motion of a particle, is known as a *Galilean transformation.* The relative velocity u is known as the parameter of the transformation. More generally we refer to the relation between the coordinates of a point as described by any two observers in relative motion with constant velocity as a Galilean transformation.

To obtain the relation between the velocities of the particle as seen by these two observers, let us differentiate both sides of (2-19). According to (2-6), the velocity v' of the particle as seen by the observer in S' is dx'/dt, whereas the velocity v as seen by the other observer is dx/dt. Hence, since u is presumed to be constant, so that $d(ut)/dt = u$, it follows that

$$v' = v - u \tag{2-20}$$

If, for example, the particle is at rest in S', then $v' = 0$ and, consistent with our expectations, (2-20) implies that $v = u$.

To illustrate the matter let us consider again the situation in Figure 2-12 and make the following identifications:

1. The system S is fixed relative to the earth.
2. The system S' is fixed relative to the train, and thus has a velocity u along the x-axis, which we take to be oriented eastward.
3. The observed velocity of the passenger in S' is v_0, and therefore

$$v' = v_0$$

Substitution of this value into (2-20) shows that the velocity v of this passenger as seen in S is

$$v = u + v_0$$

which agrees with the result obtained previously. If the passenger had been walking west, rather than east, then his observed speed on the train v' would have been $-v_0$, and for this case the observed velocity v in S would have been $v = u - v_0$.

Example 2-12 A motorboat, which cruises in still water at 12 km/hr, is cruising downstream in a current of 3 km/hr. What is its velocity relative to land, and how long will it take the boat to travel a distance of 10 km?

Solution Physically, we expect that the velocity of the boat relative to land is the sum of the velocity of the boat relative to the water plus the velocity of the water relative to land; that is, 12 km/hr + 3 km/hr = 15 km/hr. Thus the time it takes the boat to travel 10 km is

$$\frac{10\text{ km}}{15\text{ km/hr}} = 0.67\text{ hr}$$

More formally we identify S in Figure 2-13 to be stationary relative to land, and S' to be at rest relative to the water. Thus the velocity v' of the boat in S' is 12 km/hr,

and the transformation velocity u is 3 km/hr. Substituting into (2-20), we find, for the velocity v relative to land,

$$v = v' + u = (12 + 3)\ \text{km/hr} = 15\ \text{km/hr}$$

Example 2-13 A spaceship orbits the earth at a velocity, as seen from the earth, of 10 km/s. Neglecting effects due to the rotation of the earth, what is the velocity of the ship relative to the sun at an instant when the spaceship is traveling parallel to the direction of motion of the earth about its orbit?

Solution Figure 2-14 shows this situation as viewed by an observer at rest relative to the sun. The symbol v_E represents the velocity of the earth in its orbit, about the sun, which is known to be about 30 km/s, and v_s represents the unknown velocity of the ship in this coordinate system.

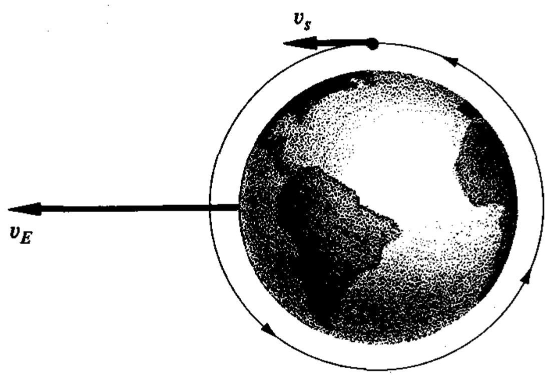


Figure 2-14

Let us select S to be the system fixed relative to the sun and S' the system at rest relative to the earth. The transformation velocity u is, for this case, v_E, and the observed velocity of the ship as seen on earth, v', is 10 km/s. Substituting these data into (2-20), we find, for $v_s = v$,

$$\begin{aligned} v_s &= u + v' = v_E + v' \\ &= 30\ \text{km/s} + 10\ \text{km/s} \\ &= 40\ \text{km/s} \end{aligned}$$

2-10 Summary of important formulas

The average velocity $\bar{v}$ of a particle that undergoes a displacement Δx during a time interval Δt is

$$\bar{v} = \frac{\Delta x}{\Delta t} \tag{2-2}$$

and its instantaneous velocity v, which is the limiting value approached by $\bar{v}$ as Δt tends to zero, is

$$v = \frac{dx}{dt} \tag{2-6}$$

The acceleration a of a particle is

$$a = \frac{dv}{dt} \tag{2-9}$$

where $v = v(t)$ is the velocity at time t.

The position $x(t)$ and the velocity $v(t)$ of a particle undergoing vertical motion in the earth's gravitational field are

$$x(t) = -\frac{1}{2}gt^2 + v_0t + h \tag{2-16}$$

$$v(t) = v_0 - gt \tag{2-17}$$

where the parameters are as defined in Figure 2-10, and $g = 9.8\ \mathrm{m/s^2}$ is the acceleration of gravity.

QUESTIONS

1. Define or describe briefly, the following terms: (a) kinematics; (b) dynamics; (c) average velocity; (d) acceleration; and (e) point particle.
2. Give an example of a position function $x(t)$ for which the average velocity $\bar{v}(t)$ is independent of the time interval Δt. Is your form for $x(t)$ unique?
3. Explain under what circumstances it is possible for the average acceleration to be independent of Δt while the average velocity varies with Δt? Illustrate by reference to a particular position function $x(t)$. Is the converse possible?
4. Under what circumstances can the earth be thought of as a point particle? Under what circumstances can it not?
5. Suppose that an automobile is driven at a velocity of 90 km/hr. If it is suddenly accelerated in the direction of motion, what happens to the velocity? What happens if it is accelerated in the opposite direction?
6. Explain why it is convenient to assume a point particle in a study of kinematics? Is it *necessary* to make this assumption? Explain.
7. What are the two possible meanings associated with each of the symbols x, y, and z in Figure 2-2?
8. What is the relation between the slope of the $x(t)$ curve for a particle and its velocity? What can you say about the velocity of a particle at a point of the $x(t)$ curve for which the slope is positive? At a point where the slope vanishes?
9. What can you say about the acceleration of a particle at a time t for which its $v(t)$ curve has a horizontal slope? What if the $v(t)$ curve has a negative slope?
10. Suppose the position function $x(t)$ for a particle is as shown in Figure 2-15. What can you say about the velocity of the particle at each of the points t_1, t_2, t_3, and t_4?

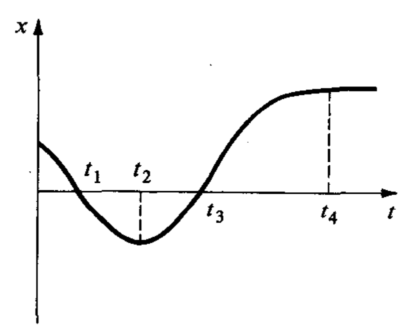


Figure 2-15

11. Repeat Question 10, but suppose this time that the curve in Figure 2-15 represents $v(t)$ as a function of t.
12. Suppose that a particle is thrown vertically upward. One of two observers standing on the ground says that its acceleration is *positive*, and the other that its acceleration is *negative.* Explain why they may both be correct! Will they both describe the initial velocity as positive?
13. An auto driver travels along a highway at 80 km/hr and observes an airplane overhead, traveling along a parallel course and in the same direction at a velocity of 250 km/hr. What is the actual velocity of the airplane with respect to the ground? What would it be if the airplane were observed to travel in the opposite direction at that speed?
14. Explain in physical terms why it is that if a boat can travel in still water at a velocity v_0, then if it travels downstream in a current of velocity u, then its velocity relative to an observer on shore is $(u + v_0)$.
15. Under what circumstances is it possible for a boat to go backward relative to a fixed point on shore?
16. Making use of the definition of a derivative, show that (2-20) is valid even though the origins of S and S' do not coincide initially. What modification of (2-19) would be necessary if initially the origins of S and S' did not coincide?

PROBLEMS

1. An automobile is first driven a distance of 50 km for 1 hr along a straight highway and then immediately afterward an additional 50 km for 45 min along the same direction. **(a)** What is the average velocity during each time interval? **(b)** What is the average velocity for the entire trip?
2. A bowling ball rolls down an alley at a uniform velocity of 5.0 m/s. **(a)** How far does it travel in 4.0 seconds? **(b)** How long does it take for the ball to travel a distance of 12 meters?
3. A stone rolls down a hill of length 30 meters. **(a)** If its time of descent is 10 seconds, what is its average velocity? **(b)** If it covers the first 15 meters in 7 seconds, what is its average velocity during the first and second half of its trip?
4. A train travels at a velocity of 80 km/hr for 30 min, and then it slows down (assume this to be instantaneous) to a speed of 50 km/hr. If the total distance traveled by the train is 100 km, how long does the trip take?
5. About how long does it take a space vehicle that is traveling at a velocity of 6000 km/hr to go from the Earth to Mars? Assume that the Earth-Mars separation distance is 8.0×10^7 km.
6. How long in years would it take an astronaut traveling at a velocity of 6.0×10^4 km/hr to go to Alpha Centauri, which is a star about 4.0×10^{13} km from us?
7. An automobile starting from rest achieves a velocity of 2.0 m/s in a time interval of 3.0 seconds. **(a)** What is the average acceleration? **(b)** What would the average acceleration have been if the automobile had achieved this speed in half the time?
8. A rocket takes off from the surface of the earth and achieves a height of 10 km in a time interval of 10 seconds. **(a)** What is the average veloc-

ity? (b) What is the average acceleration?

9. A boat starts from rest and travels in a straight line across a lake.
   (a) What is its average velocity during a 15-min interval in which it covers 2 km?
   (b) What is its average acceleration if it starts from rest and achieves a velocity of 10 km/hr in 15 min?
   (c) If at the end of 30 min it is traveling at a velocity of 8 km/hr, what is its average acceleration during this second 15-min interval? Interpret the negative sign in your answer.
   (d) What is the average acceleration over the entire 30-min trip?
10. With t in seconds and x in meters, the position function for a particle is

$$x(t) = 3t^2 + 2t + 4$$

(a) What is the position of the particle at $t = 1$ second? At $t = 3$ seconds? (b) What is the total distance traveled in the time interval 1 second $\leq t \leq 2$ seconds? (c) What is the average velocity during the time interval in (b)?

11. For the position function in Problem 10, calculate the average velocity of the particle at time t over a subsequent time interval Δt. Assuming that $t = 1$ second, calculate $\bar{v}$ for $\Delta t = 1$ second, 0.1 second, 10^{-5} second, and guess the value for the instantaneous velocity at this time, $t = 1$ second.
12. Calculate the velocity at time t for the particles whose position functions in SI units are (a) $x = -17t^2$; (b) $x = 3 - 5t$; (c) $x = 22t^2 - 4$; (d) $x = 2t - 4t^3$; (e) $x = 6t^7 - 4t^4$.
13. Assuming that x_0, a, v_0, α, and β are fixed parameters, calculate the velocity at time t associated with each of the following position functions:
   (a) $x = \alpha t^4 - \beta t^3$;
   (b) $x = -\frac{1}{2}at^2 + v_0t$;
   (c) $x = x_0 - 3t^2$;
   (d) $x = x_0 + \alpha t^4$;
   (e) $x = (\alpha/2)t^4 - (\beta/3)t^3 - at^2 + v_0t - x_0$.
14. State the dimensions, in SI units, of the various parameters in each part of Problem 13.
15. For each of the following position functions calculate the velocity of the particle at those instants when it travels through the origin of the coordinate axis. Assume the t is measured in seconds, and x in meters. (a) $x = 3t$; (b) $x = 4t^2 - 2t$; (c) $x = 2t - 8t^3$.
16. Calculate the position function $x(t)$ of a particle at time t if its velocity $v(t)$ is given by the following formulas. Assume that $v(t)$ is expressed in meters per second and t is in seconds and that at $t = 0$, the particle is located at the point indicated in parentheses after the velocity formula. (a) $v(t) = -3$, (2); (b) $v(t) = -16t + 3$, (0); (c) $v = 4t + 7$, (−4); (d) $v = 16t^2 - 4t$, (5); (e) $v = (5 + t)^2$, (3).
17. Calculate the acceleration for each of the position functions in Problem 12.
18. Calculate the acceleration of the particle for each of the position functions in Problem 13.
19. For each of the following acceleration formulas, calculate the position function of the particle at time t. Assume in each case that the initial position and velocity are given in parentheses to the right of the formula, with the first number the velocity, and that SI units are being used. (a) $a(t) = -2$, (0, 0); (b) $a(t) = -9.8$, (0, 2); (c) $a(t) = -3t$, (1, 1); (d) $a(t) = -3t^2 + 2t + 1$, (2, 2).
20. An automobile starts from rest and accelerates nonuniformly in accordance with the formula $a(t) = bt$.

(a) If the units of $a(t)$ are meters per second squared, what are the units for the constant b?
(b) Show that the velocity is proportional to t^2 and find the coefficient of proportionality.
(c) How far in terms of b will the automobile travel during the initial 10 seconds?

21. A positively charged pion (π^+) is traveling at a velocity of 2000 m/s when it enters the region between two capacitor plates where it experiences a uniform acceleration of $5.0 \times 10^8\ \text{m/s}^2$ directed opposite to the original direction of motion of the pion.
(a) How long does it take the π^+ to come to rest?
(b) How far has it traveled in this time?
(c) How long does it take to return to its starting point and what is its velocity on arriving there?

22. A baseball is thrown vertically upward with an initial velocity of 20 m/s. (a) How high does it rise? (b) How long does it take to get to the top of its trajectory?

23. How long after being thrown does the baseball of Problem 22 return to its starting point? What is its velocity at this instant?

24. An apple is dropped out of a window 25 meters above the ground. How far below its launching point is it after 1 second? after 2 seconds? What is its velocity at these two instants?

25. How long after being released will the apple of Problem 24 strike the ground? What is its velocity on impact?

26. Show that if an object is dropped from any height, then the distances traversed in equal intervals of time are in the same ratios as the odd integers. (*Note*: This result was first deduced by Galileo.)

27. A bullet is fired vertically upward from the ground with a muzzle velocity of 100 m/s.
(a) How high does the bullet rise?
(b) How long after being fired will it strike the ground?
(c) What is the velocity of the bullet at an instant when it has achieved half of its maximum height?

28. Derive (2-18) by explicitly eliminating the time between (2-16) and (2-17).

29. By multiplying both sides of (2-15) by v, show by use of (A-4) that the result may be expressed as

$$\frac{d}{dt}(v^2 + 2gx) = 0$$

and thus derive (2-18) in this way.

30. Show in general that if a particle undergoes a constant acceleration, then its average velocity $\bar{v}(t)$ over a time interval Δt is

$$\bar{v}(t) = \frac{1}{2}[v(t + \Delta t) + v(t)]$$

where $v(t + \Delta t)$ and $v(t)$ are the instantaneous velocities at the times $(t + \Delta t)$ and t, respectively.

31. On a certain planet, a rock is launched vertically upward with a velocity of 10 m/s. It is observed to return to its starting point in a time interval of 5.0 seconds. (a) What is the value of "g" on this planet? (b) How high did the rock rise?

32. An astronaut on the moon throws a rock vertically upward and finds that it returns to its starting point in 22 seconds. The maximum height reached by the rock is 100 meters. (a) What is the value of "g" on the moon? (b) What is the initial velocity with which the rock was thrown.

33. If a particle is thrown upward from the ground and reaches a maximum height H, show that its initial veloc-

ity v_0 must be given by

$$v_0 = \sqrt{2gH}$$

34. An object is thrown vertically upward from a point 10 meters above the ground with a velocity of 5.0 m/s. **(a)** What maximum height is achieved? **(b)** How long after being launched will it strike the ground? **(c)** What is its velocity at impact?

***35.** A small marble is hurled vertically downward from a height of 100 meters with an initial *downward* velocity of 2 m/s. Precisely 1.0 second later, a second marble is thrown downward with an initial speed v_0 from the same place.

(a) At what time does the first marble strike the ground?

(b) What must be the value of v_0 so that the second marble strikes the first at the moment of impact with the ground?

(c) What are the velocities of the marbles at the instant of collision in (b)?

†36. A man is in an elevator, which is descending at a uniform velocity of 2.0 m/s, when he drops a pellet.

(a) Show that the velocity $v(t)$ of the pellet as seen by a *stationary* observer is $v(t) = -9.8t - 2$, where $v(t)$ is in meters per second and t is in seconds.

(b) What is the observed velocity v' of the particle as seen by the man in the elevator?

(c) Calculate the position of the pellet as seen by the man in the elevator if the pellet is dropped from a point 3.0 meters above the floor of the elevator.

†37. A particle travels along a straight line under the influence of an acceleration $a(t)$, which has the form

$$a(t) = -bt^2$$

with $a(t)$ in meters per second squared and t in seconds.

(a) Calculate the position at any time t of this particle in a coordinate system S in which at $t = 0$ the particle is at the origin and has an initial velocity $+3$ m/s.

(b) Repeat (a), but this time in a coordinate system S', where again at $t = 0$ the particle is at the origin but where at $t = 1$ second it is at the point $x' = 3$ meters.

(c) Calculate the relative velocity between the S and S' systems in (a) and (b).

3 Two-dimensional kinematics

Geometry remains a mathematical science because the deduction of theorems from axioms remains a purely logical problem; at the same time it is a physical science insofar as its axioms contain assertions relating to natural objects the validity of which can be proved only by experience.

ALBERT EINSTEIN

3-1 Introduction

The purpose of this chapter is to generalize the discussion of one-dimensional motion in Chapter 2 to the physically more interesting case of a particle traveling along an arbitrary path through space. For simplicity, the discussion will be restricted to those situations for which the trajectory of the particle lies entirely in a plane. In effect, therefore, we shall be studying the kinematics of a particle in two dimensions. The generalizations required to treat the three-dimensional case are relatively straightforward and the essential features can be understood without going into many of the geometrical complications.

The description of two- and three-dimensional kinematics is most conveniently carried out by making use of certain physical quantities known as *vectors*. Accordingly, the first several sections of this chapter are devoted to a discussion of vectors and some of their important properties. Using vectors as a tool, we then turn in the remainder of the chapter to a discussion of the kinematics of a particle moving in a plane.

3-2 Vectors and their representations

For a variety of purposes, it is convenient to classify physical quantities into the two classes of *scalars* and *vectors.* A scalar is a physical quantity that is completely specified by a numerical value and a unit. In particular, a scalar does *not* have associated with it any sense of direction. The mass of a proton, the number of molecules in a glass of water, and the volume or temperature of a room are all scalars. By contrast, a physical quantity whose specification involves not only a numerical value and a unit but a sense of direction as well is called a *vector.* Some examples of vector quantities are velocity, acceleration, force, and magnetic field strength—each of which has associated with it not only a magnitude but a sense of direction as well.

In order to distinguish between vectors and scalars, **boldface type** will be used to designate the symbols for vectors and ordinary *italics* will be used for scalars. Thus we shall write **A** and **B** to represent the two vectors **A** and **B**. Sometimes we are interested only in the magnitude of a vector and not in its direction and we shall represent magnitudes also by use of *italics.* Thus the two symbols A and B represent the (positive) magnitudes of the vectors **A** and **B**, respectively. Another way for expressing the magnitude of a vector is by placing a vertical bar (an absolute value sign) on each side of the symbol for the vector. Thus the symbols $|\mathbf{A}|$ and $|\mathbf{B}|$ will also represent the magnitudes A and B of the vectors **A** and **B**, respectively. Note that the magnitude of a vector is a strictly non-negative quantity and since by definition it involves no sense of direction, it is itself a scalar. A vector whose magnitude vanishes is called a *null vector.*

A convenient way to represent vectors geometrically is by use of arrows (see Figure 3-1). Given the magnitude and direction of a vector, we select some convenient unit of length and draw an arrow that points along the direction of the vector and with length proportional to its magnitude. Thus, suppose the vector of interest is velocity and that one unit of length represents a velocity of magnitude of 10 km/s. Then if the vector **A** in Figure 3-1 is two units long and thus represents a velocity of 20 km/s, due north, then **B**, which is only one unit in length, represents a velocity of 10 km/s, due east. Similarly **C** represents a velocity of 30 km/s northeast, and **D**, which is equal and opposite to **C**, represents a velocity of 30 km/s, directed toward the southwest.

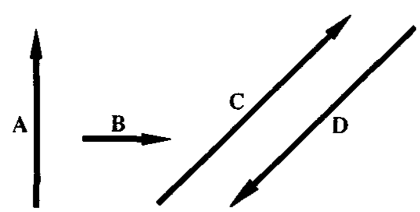


Figure 3-1

It follows from this discussion that only the length of the arrow and its orientation, but not its precise location, are relevant when vectors are represented in this way. Thus the arrow representing any vector may be moved about at pleasure provided only that the direction in which it points is not changed. Figure 3-2 shows three arrows representing the vectors **A**, **A**′, **A**″. Since the three arrows are all of the same length and point in the same direction, it follows that they all represent precisely the same vector.

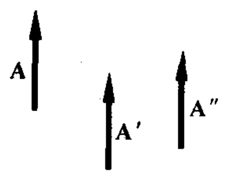


Figure 3-2

3-3 The displacement and position vectors

In a study of both kinematics and dynamics, one of the very important vectors that arises is associated with the change in the position of a particle and is known as the *displacement vector.* The purpose of this section is to define this vector and the related one known as the *position vector.*

Suppose in Figure 3-3a a particle, which is originally at the point O, is taken to a neighboring point P along an arbitrary path such as the one represented by the dashed line in the figure. The *displacement* **d** of the particle is defined to be its net change in position during this process and is represented by the arrow drawn from O to P. Note that the displacement vector **d** represents the net change in the position of the particle and is by definition the same regardless of the path along which the particle is transported from O to P. For example, if the particle is taken along the dotted path shown in Figure 3-3b, its displacement is still **d**, since its initial and final positions are still O and P, respectively. The magnitude $d = |\mathbf{d}|$ of the displacement vector **d** represents the distance from O to P; that is, d is the length of the shortest path connecting O to P.

Figure 3-4 shows two particular paths along which a particle can go from O to P. In the upper one of these, it goes along the straight-line path from O

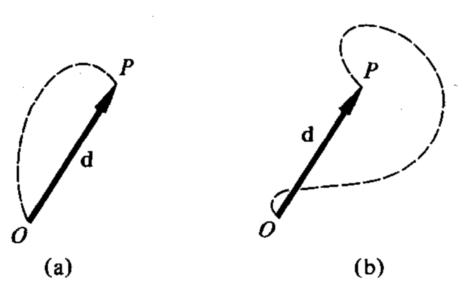


Figure 3-3

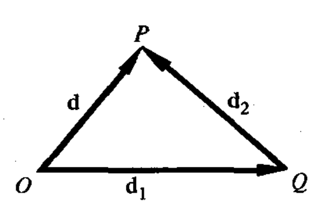


Figure 3-4

to P, whereas along the lower path it first undergoes the displacement $\mathbf{d}_1$ to the point Q and then a second displacement $\mathbf{d}_2$ from the point Q to the final position P. The net displacement along both paths is the same, namely $\mathbf{d}$ since the particle originates on O and ends up at P in both cases. This suggests that the addition of two displacement vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ be defined by the formula

$$\mathbf{d} = \mathbf{d}_1 + \mathbf{d}_2 \tag{3-1}$$

provided that $\mathbf{d}$, $\mathbf{d}_1$, and $\mathbf{d}_2$ form the sides of triangle as in Figure 3-4. Indeed, in Section 3-4 vector addition will be defined in precisely this geometric way.

Figure 3-5 shows how the position of a particle may be represented by a displacement vector. Consider a particle located at a point P and construct a coordinate system with origin at the point O. We define the *position vector* $\mathbf{r}$ of the particle to be its displacement from O. Thus, as shown in the figure, the position vector $\mathbf{r}$ for the particle may be represented by an arrow drawn from the origin of the coordinate system to the position P. It should be emphasized that the particle has *not* been taken from O to P, but rather it is *at P* and its position vector $\mathbf{r}$ is defined to be its displacement vector *if it had been at O and were moved to P.* The magnitude $r = |\mathbf{r}|$ of $\mathbf{r}$ represents the separation distance between the particle and the origin, and the direction of $\mathbf{r}$ is the direction along the straight line it is necessary to travel in order to reach it from O. Note from the figure that only the location of the origin plays a role in the definition for $\mathbf{r}$; the orientation of the x- and y-axes is entirely inconsequential.

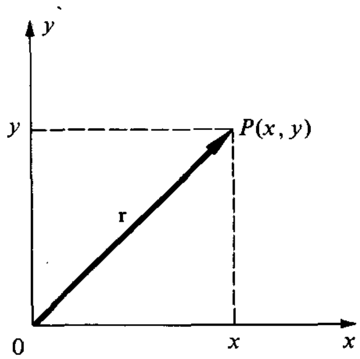


Figure 3-5

3-4 Vector algebra

In a restricted sense, many of the operations of addition, subtraction, and multiplication that we perform with scalars can also be carried out with vectors. The purpose of this section is to define some of these operations involving vectors.

The product of a scalar and a vector

Consider an arbitrary vector **A** and a *positive* scalar λ. We define their product $\lambda\mathbf{A}$ to be itself a vector which points in the same direction as does **A** and has a magnitude $\lambda|\mathbf{A}|$. Figure 3-6 shows three vectors **A**, $\frac{1}{2}\mathbf{A}$, and $2\mathbf{A}$. Note that, consistent with this definition, the vectors $\frac{1}{2}\mathbf{A}$ and $2\mathbf{A}$ both point in the same direction as does **A**, but with the respective magnitudes of $\frac{1}{2}$ and 2 times that of **A** itself. If, for example, **A** represents a velocity of 90 km/hr, due east, then $\frac{1}{2}\mathbf{A}$ represents a velocity of $\frac{1}{2}\times 90$ km/hr $= 45$ km/hr, also directed eastward.

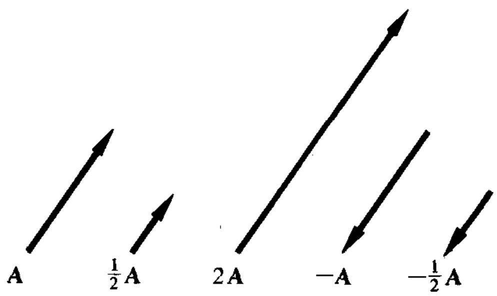


Figure 3-6

For the case that the scalar parameter λ is negative, the product $\lambda\mathbf{A}$ is defined to be again a vector with magnitude $|\lambda|$ times that of **A** but to be oriented in a direction *opposite* to **A**. Figure 3-6 also shows the vectors $-\mathbf{A}$ and $-\frac{1}{2}\mathbf{A}$. Note that, consistent with our definition, **A** and $-\mathbf{A}$ have equal magnitudes but opposite orientations and that the magnitude of the vector $-\frac{1}{2}\mathbf{A}$ is one half that of **A**, but its direction is opposite to the direction of **A**. For example, if **A** represents a displacement of 2 km, due north, then $-3\mathbf{A}$ represents a displacement of 6 km, due south.

Associated with any vector **A** is a unit vector $\mathbf{u}_A$, which is defined to be a vector that points in the same direction as does **A**, but has the magnitude unity. An explicit formula for $\mathbf{u}_A$ is given by

$$\mathbf{u}_A = \frac{1}{A}\mathbf{A} \tag{3-2}$$

for according to the above definition, the quantity $\mathbf{A}/A$ is oriented along **A** and has magnitude unity, since

$$|\mathbf{u}_A| = \frac{1}{A}|\mathbf{A}| = 1$$

The addition of vectors

Given two vectors **A** and **B**, as in Figure 3-7a, we define their sum $(\mathbf{A}+\mathbf{B})$ in accordance with the construction shown in Figure 3-7b. Specifically, making

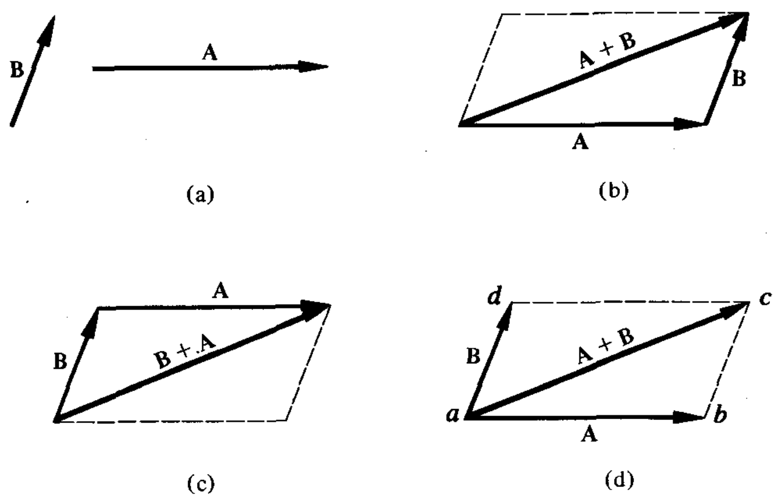


Figure 3-7

use of the fact that the arrow representing a vector can be moved about freely, we take the arrow representing **B** and move it—always so that it continues to be oriented parallel to its original direction—until its tail touches the tip of **A**. The arrow representing the vector sum $(\mathbf{A}+\mathbf{B})$ is then found by drawing a directed line from the tail of **A** to the tip of **B**. Note that the definition for vector addition is precisely that anticipated in Figure 3-4 for the displacement vector for a particle.

If we apply this definition of vector addition to **A** and **B** but this time in the opposite order (that is, by keeping **B** fixed and moving **A**), then as shown in Figure 3-7c we obtain the vector $(\mathbf{B}+\mathbf{A})$. It is apparent from the geometry of these constructions that the process of vector addition is *commutative.* That is,

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A} \tag{3-3}$$

so in carrying out vector addition it is immaterial in which order the two vectors are added. Finally, in Figure 3-7d this commutativity property is used to obtain an alternate way for obtaining the sum of two vectors graphically. Here they are placed so that they form the adjacent sides of a parallelogram *abcd.* The sum $(\mathbf{A}+\mathbf{B})$ is then a vector directed as shown in the figure and with magnitude numerically equal to the length of the diagonal *ac*. The vector sum $(\mathbf{A}+\mathbf{B})$ is also known as the *resultant* of **A** and **B**.

A second important property of vector addition is that it is *associative.* This means that if we wish to add together three vectors **A**, **B**, and **C**, then it is immaterial whether we first form the sum $(\mathbf{A}+\mathbf{B})$ and then add **C** to its resultant, or whether we first find the resultant of **B** and **C** and then add **A** to it. In mathematical terms this *associativity* property is expressed by

$$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C}) \tag{3-4}$$

and its validity is established in Figure 3-8.

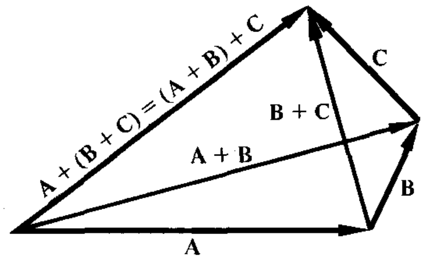


Figure 3-8

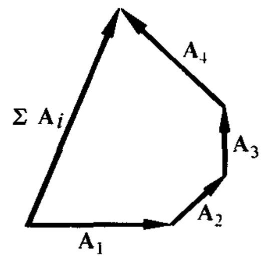


Figure 3-9

More generally, to find the vector sum of any number of vectors we arrange the arrows representing the vectors in some order and move them around successively until the tail of each one is in contact with the tip of its precursor in this ordering. The arrow drawn from the tail of the first to the tip of the last then represents the desired vector sum. Figure 3-9 illustrates this process for the vector sum $\Sigma\mathbf{A}_i = \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 + \mathbf{A}_4$. Because of (3-3) and (3-4), the order in which the vectors are arranged is not relevant. For the special case that the tip of the last vector touches the tail of the first, so that the vectors form a closed polygon, it follows that the vector sum vanishes! In other words, the vector sum of any number of vectors that form a closed polygon is the null vector.

Vector subtraction

The vector difference $(\mathbf{A} - \mathbf{B})$ of two vectors $\mathbf{A}$ and $\mathbf{B}$ is defined to be the sum of the vectors $\mathbf{A}$ and $(-\mathbf{B})$. That is,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) \tag{3-5}$$

and this is illustrated graphically in Figure 3-10. As shown, with $\mathbf{A}$ fixed, $\mathbf{B}$ is moved until its tail touches that of $\mathbf{A}$. The vector $(\mathbf{A} - \mathbf{B})$ is then found by drawing an arrow from the tail of $\mathbf{A}$ to the tip of $-\mathbf{B}$.

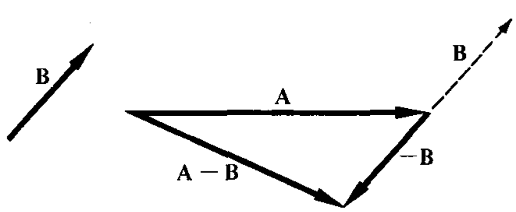


Figure 3-10

Example 3-1 An airplane flies due north for 75 km and then west for an additional 100 km. What is the displacement of the airplane?

Solution If $\mathbf{A}$ represents the northward displacement, and $\mathbf{B}$ the corresponding displacement westward, then as shown in Figure 3-11 the net displacement is given by the vector sum $(\mathbf{A} + \mathbf{B})$. Making use of the fact that $\mathbf{A}$ and $\mathbf{B}$ are at right angles, by use of

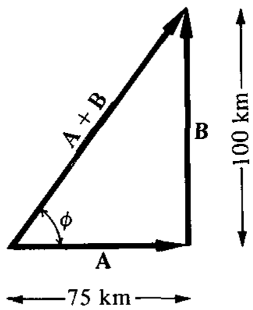


Figure 3-11

the Pythagorean theorem we obtain

$$|\mathbf{A}+\mathbf{B}| = [(75\text{ km})^2 + (100\text{ km})^2]^{1/2} = 125\text{ km}$$

Further, since the triangle is a right triangle, the sine of the angle ϕ that the vector $(\mathbf{A}+\mathbf{B})$ makes with $\mathbf{A}$ is given by

$$\sin\phi = \frac{100\text{ km}}{125\text{ km}} = \frac{4}{5}$$

or

$$\phi \simeq 53°$$

3-5 Components of a vector

Consider a vector **A** and a directed line—such as a coordinate axis—oriented along the direction of a vector **u**. The *component* A_u of the vector **A** along the direction of **u** is defined to be the projection of **A** onto this direction. If, as in Figure 3-12a, **A** is moved until its tail is in contact with that of **u**, then the projection of **A** onto **u** is represented by the part of the line between this point and the intersection of **u** with the perpendicular from the tip of the vector **A** onto **u**. In terms of the angle θ between **A** and **u**, the component A_u may be expressed by

$$A_u = A\cos\theta \tag{3-6}$$

Note that, depending on the value of θ, A_u may be positive or negative or zero. Figure 3-12b shows a case for which the angle θ is obtuse and for which therefore the component of the vector **B** is negative.

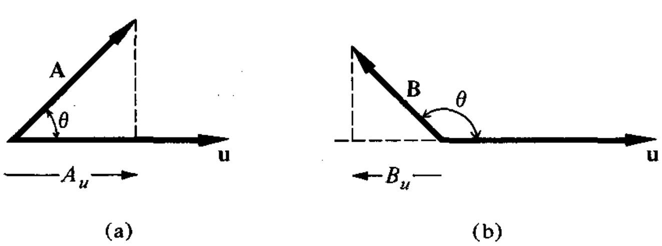


Figure 3-12

Consider in Figure 3-13 a vector **A** with its tail at the origin of a certain Cartesian coordinate system. If θ is the angle between **A** and the positive x-axis, then the components A_x and A_y of the vector **A** along the x- and y-axes, respectively, are

$$A_x = A \cos\theta \qquad A_y = A \sin\theta \tag{3-7}$$

where the second equality follows from (3-6) and the fact that $\cos(90° - \theta) = \sin\theta$. Although it has been assumed in Figure 3-13 that θ is an acute angle, it is easily confirmed that (3-7) is valid for any value of θ. Thus, given any two-dimensional coordinate system, any vector in its plane can be expressed in terms of its components along the two coordinate axes. In particular, if the vector under discussion is the position vector **r** of the particle in Figure 3-5, then its x-coordinate is the x-component of **r**, and similarly the y-coordinate is the y-component of **r**.

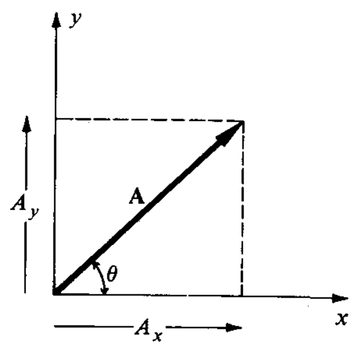


Figure 3-13

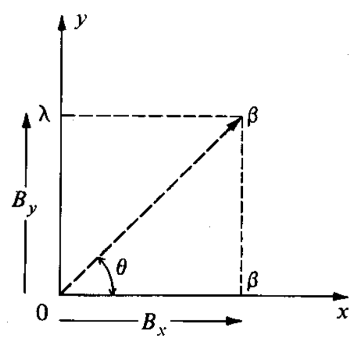


Figure 3-14

Let us now examine the inverse problem. Given a coordinate system, as in Figure 3-14, and the components B_x and B_y of a vector **B** along its two axes, is it possible to reconstruct the vector **B** from these data? The answer is yes. For if we draw a horizontal line $\lambda\beta$ a distance B_y above (or below, if $B_y < 0$) the x-axis, and a corresponding vertical line $\beta\beta$ a distance B_x to the right of the y-axis, then an arrow drawn from the origin to the intersection of these two lines will be a vector whose components along the two axes will be precisely the given values B_x and B_y. Hence it is the sought-for vector **B**. Making use of the Pythagorean theorem and the laws of trigonometry, we may express the magnitude and the direction of the vector **B** by the formulas

$$B = \sqrt{B_x^2 + B_y^2} \tag{3-8}$$

and

$$\theta = \tan^{-1}\frac{B_y}{B_x} \tag{3-9}$$

Thus, given the components B_x and B_y of a vector **B** along two mutually orthogonal axes, the vector itself is uniquely given in (3-8) and (3-9). Although it has been assumed in this discussion that B_x and B_y are both positive, (3-8) and (3-9) are valid regardless of their signs.

Figure 3-15 gives a geometric proof of the statement that if A_x, A_y, and B_x, B_y are the components of the two vectors **A** and **B** along the axes of a certain coordinate system, then the components of the vector $(\mathbf{A}+\mathbf{B})$ along these same axes are $(A_x + B_x)$ and $(A_y + B_y)$. Similarly, in the problems it is established that the components of $\lambda\,\mathbf{A}$ are λA_x and λA_y, and the components of the vector $(\mathbf{A}-\mathbf{B})$ are $(A_x - B_x)$ and $(A_y - B_y)$.

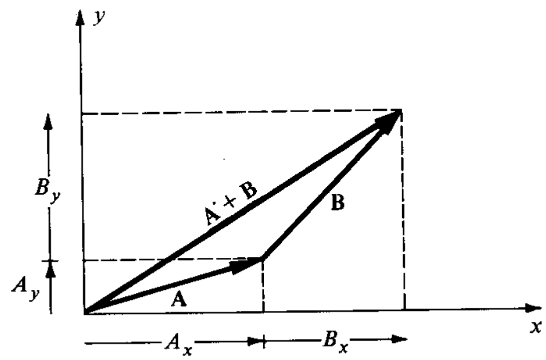


Figure 3-15

Example 3-2 An automobile while traveling along a straight highway, which is directed 60° north of east, covers a distance of 90 km in 1 hr. How far east and north of its starting point has the auto been displaced in this time?

Solution As in Figure 3-13, let us set up a coordinate system, with origin at the original position of the auto and with the y-axis to the north, so that the x-axis points toward the east. From the given data we know that the magnitude of the displacement vector **A** is 90 km and that the angle it makes with the x-axis is 60°. The components A_x and A_y represent the distances that the auto has been displaced toward the east and north, respectively. Applying (3-7), we find that

$$A_x = A \cos\theta = 90\text{ km}\cos 60° = 45\text{ km}$$
$$A_y = A \sin\theta = 90\text{ km}\sin 60° = 78\text{ km}$$

since $\cos 60° = \frac{1}{2}$, and $\sin 60° = \frac{1}{2}\sqrt{3}$.

Example 3-3 An airplane is cruising along a certain direction. If after flying for 1 hr the pilot finds himself 300 km east and 600 km south of his starting point, what is his actual vectorial displacement during this time interval?

Solution Let us set up a system of coordinate axes as in Figure 3-16, with the x-axis directed eastward. This time we are given the components d_x and d_y of the displacement vector **d**,

$$d_x = 300\text{ km}, \qquad d_y = -600\text{ km}$$

where the minus sign is required since the positive y-axis is directed northward. Substituting these values into (3-8), for the magnitude of **d** we obtain

$$d = \sqrt{d_x^2 + d_y^2} = [(300\text{ km})^2 + (600\text{ km})^2]^{1/2}$$
$$= 671\text{ km}$$

Similarly, in terms of the angle θ in the figure, we have

$$\theta = \tan^{-1}\frac{600}{300} = \tan^{-1} 2 \cong 63°$$

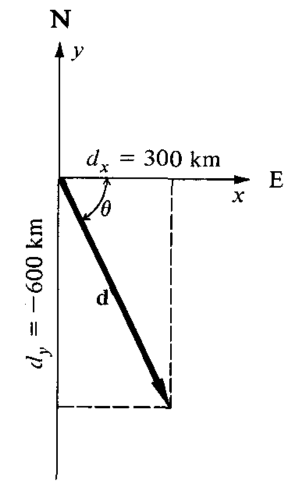


Figure 3-16

3-6 Basis vectors

It is frequently convenient to be able to express a vector directly in terms of its components in a given coordinate system. The purpose of this section is to show how this can be achieved by the use of *basis vectors.*

Consider in Figure 3-17a a set of coordinate axes, and let **i** and **j** represent *unit* vectors along the positive senses of the x- and y-axes, respectively. These two vectors **i** and **j** constitute the *basis vectors* for the given coordinate system. By definition, a unit vector has magnitude unity. Hence for any basis vectors **i** and **j**, we have

$$|\mathbf{i}| = |\mathbf{j}| = 1 \tag{3-10}$$

Suppose now, as shown in Figure 3-17b, that in this same coordinate system in which **i** and **j** are basis vectors the components of a certain vector **A** are A_x and A_y. Let us construct, as in Figure 3-17c, the two vectors $\mathbf{i}A_x$ and $\mathbf{j}A_y$. Since **i** and **j** are unit vectors, it follows that the former of these, $\mathbf{i}A_x$, is a

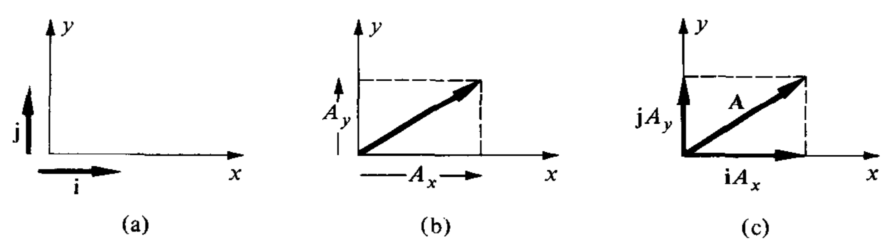


Figure 3-17

vector with component A_x along the x-axis and zero component along the y-axis, whereas the other, $\mathbf{j}A_y$, has a vanishing component along the x-axis and the component A_y along the y-axis. It follows that the vector sum $(\mathbf{i}A_x + \mathbf{j}A_y)$ represents a vector with components A_x and A_y along the x- and y-axes, respectively, and therefore it must be the vector $\mathbf{A}$ itself. Thus the vector $\mathbf{A}$ may be represented by

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y$$

More generally,

If B_x, B_y are the components of a vector $\mathbf{B}$ in a fixed coordinate system, then

$$\mathbf{B} = \mathbf{i}B_x + \mathbf{j}B_y \tag{3-11}$$

where $\mathbf{i}$ and $\mathbf{j}$ are the basis vectors and are parallel, respectively, to the x- and y-axes of that system.

In particular, if the instantaneous coordinates of a particle moving in a plane are (x, y), then its position vector $\mathbf{r}$ in this system may be expressed in the form

$$\mathbf{r} = \mathbf{i}x + \mathbf{j}y \tag{3-12}$$

since, as we saw previously, the components of the position vector $\mathbf{r}$ in any coordinate system are precisely the coordinates of the particle in that system.

In Figure 3-15 it was established geometrically that the components of the sum of two vectors is the sum of their components. Making use of the above ideas of representation of vectors, we can now give a simple algebraic proof of this statement. If $\mathbf{A}$ and $\mathbf{B}$ are any two vectors with the respective components A_x, A_y and B_x, B_y in a given system, and if $\mathbf{i}$ and $\mathbf{j}$ are basis vectors in this system, then

$$\begin{aligned}\mathbf{A} + \mathbf{B} &= (\mathbf{i}A_x + \mathbf{j}A_y) + (\mathbf{i}B_x + \mathbf{j}B_y)\\ &= \mathbf{i}(A_x + B_x) + \mathbf{j}(A_y + B_y)\end{aligned}$$

The second equality here follows by use of the associativity property of vector addition in (3-4).

3-7 Vector functions

For the remainder of this chapter we shall be concerned with the kinematical problem of describing the motion of a particle in a plane.

Consider in Figure 3-18 a particle that is moving along a general two-dimensional trajectory AB, and suppose that at time t it is at the point P with

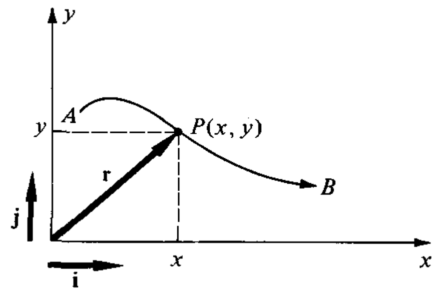


Figure 3-18

the coordinates (x, y). As the particle travels along AB, in general both its x-coordinate and its y-coordinate will change. Thus in order to give a complete description for the motion of the particle, it is necessary to specify *two* functions of time: the x-coordinate $x(t)$ of the particle at time t and the y-coordinate $y(t)$ at time t. In other words, the analogue of the single function of time $x(t)$, which in one dimension describes the motion of a particle confined to the x-axis, is the ordered pair of functions $[x(t), y(t)]$, whose complete specification is equivalent to a knowledge of the whereabouts of the particle at any time t.

Equivalently, the position of the particle may be described by the position vector

$$\mathbf{r}(t) = \mathbf{i}x(t) + \mathbf{j}y(t) \tag{3-13}$$

where, as shown in Figure 3-18, $\mathbf{i}$ and $\mathbf{j}$ are the basis vectors of the underlying coordinate system. As the particle travels along AB, this position vector $\mathbf{r}(t)$ changes both its magnitude and direction but in such a way that at each instant the tip of the arrow representing $\mathbf{r}(t)$ coincides with the instantaneous position of the particle. It is significant to note that this way of thinking and of describing the motion of the particle is independent of the choice of coordinate system. It is thus fully consistent with the physical fact that the motion of a particle has nothing whatsoever to do with the coordinate system in terms of which this motion is described. A coordinate system is simply a convenient way for describing the position of a particle.

A vector of the type just described, which depends on an independent variable and changes as this variable does, is known as a *vector function*.

3-8 Velocity

Consider, in Figure 3-19, a particle moving along a trajectory AB and suppose that at the instant t it is at P_1 and that a small time interval Δt later it is at P_2. According to the definition of a displacement vector, the arrow drawn from P_1 to P_2 represents the displacement $\mathbf{d}$ for the particle during this time interval Δt. By analogy to our discussion of one-dimensional kinematics, we define the

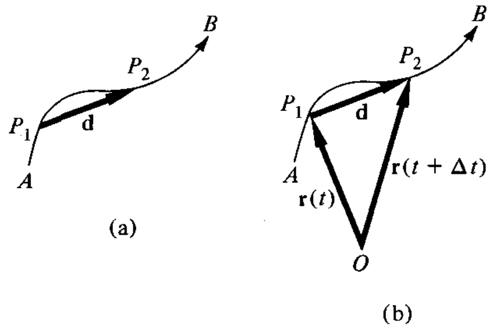


Figure 3-19

average velocity $\bar{\mathbf{v}}(t)$ over the time interval Δt to be the vectorial displacement of the particle per unit time:

$$\bar{\mathbf{v}}(t) = \frac{\mathbf{d}}{\Delta t} \tag{3-14}$$

Note that, in the present case, $\bar{\mathbf{v}}(t)$ is a vector quantity and is directed along $\mathbf{d}$. For the special case of motion along a straight line, it is easy to confirm that (3-14) is equivalent to the definition in (2-3) for one-dimensional motion.

An equivalent way of expressing $\bar{\mathbf{v}}(t)$ may be obtained in the following manner. As in Figure 3-19b, let us fix an origin O at some point near the trajectory AB and define the position vector $\mathbf{r}(t)$ of the particle relative to this origin. If $\mathbf{r}(t)$ represents the position of the particle when it is at P_1, then $\mathbf{r}(t+\Delta t)$ will be the corresponding position vector when it is at P_2. Since the vectors $\mathbf{r}(t+\Delta t)$, $\mathbf{r}(t)$, and $\mathbf{d}$ form a closed triangle, it follows from the definition of vector addition that

$$\mathbf{d} = \mathbf{r}(t+\Delta t) - \mathbf{r}(t)$$

Substituting this form for $\mathbf{d}$ into (3-14) we find for the average velocity $\bar{\mathbf{v}}(t)$ the equivalent form

$$\bar{\mathbf{v}}(t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t)}{\Delta t} \tag{3-15}$$

The similarity between this formula and (2-3) should be noted.

As before, the average velocity $\bar{\mathbf{v}}(t)$ as here defined depends on the length of the time interval Δt, and in general as Δt varies so does $\bar{\mathbf{v}}(t)$. The *instantaneous velocity*, or the *velocity* $\mathbf{v}(t)$, is defined to be independent of Δt and to be the limiting value achieved by $\bar{\mathbf{v}}$ as Δt tends to zero. Thus

$$\begin{aligned}\mathbf{v}(t) &= \lim_{\Delta t \to 0} \bar{\mathbf{v}}(t) \\ &= \lim_{\Delta t \to 0} \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t)}{\Delta t}\end{aligned} \tag{3-16}$$

where the second equality follows by use of (3-15). As for scalar functions, the limiting form here is by definition the derivative of $\mathbf{r}(t)$ with respect to time:

$$\mathbf{v}(t) = \frac{d}{dt}\mathbf{r}(t) \tag{3-17}$$

If we express $\mathbf{r}(t)$ in terms of the coordinates $[x(t), y(t)]$ of the particle by use of (3-13), then by the use of (A-2) and the fact that basis vectors are independent of time we obtain,

$$\mathbf{v}(t) = \frac{d}{dt}[\mathbf{i}x(t) + \mathbf{j}y(t)]$$

$$= \mathbf{i}\frac{dx}{dt} + \mathbf{j}\frac{dy}{dt} \tag{3-18}$$

This states that if in a given coordinate system the trajectory of a particle in a plane is described by its two components $x(t)$ and $y(t)$, then its observed velocity $\mathbf{v}(t)$ is given by a vector that in the same coordinate system has the components v_x, v_y:

$$v_x = \frac{dx}{dt} \qquad v_y = \frac{dy}{dt} \tag{3-19}$$

It is interesting to note that although the position vector $\mathbf{r}(t)$ depends on the choice of origin O, the velocity $\mathbf{v}$ of the particle does not. For if we change the origin O in Figure 3-19b, then even though both of the vectors $\mathbf{r}(t)$ and $\mathbf{r}(t + \Delta t)$ will separately change, their difference $\mathbf{d}$ remains unaltered. Thus, since $\mathbf{v}$ is defined only in terms of this displacement vector $\mathbf{d}$, it follows that the velocity $\mathbf{v}$ is invariant to this choice of origin.

The magnitude d of the displacement vector $\mathbf{d}$ of a particle has been defined as the distance between its initial and final positions. Similarly, the *speed* of a particle is defined as the *magnitude of its velocity*. Thus we say that an airplane that travels at a *velocity* of 900 km/hr, due east, has a speed of 900 km/hr.

Example 3-4 Suppose that in a certain coordinate system the position vector for a particle is

$$\mathbf{r}(t) = at\mathbf{i} - bt^2\mathbf{j}$$

where $a = 2.0$ m/s and $b = 2.0$ m/s^2. Calculate the velocity of the particle at $t = 0$ and at $t = 1$ second, and its speed at these two instants.

Solution Substituting the forms

$$x(t) = at \qquad y(t) = -bt^2$$

into (3-19), we obtain

$$v_x = \frac{d}{dt}(at) = a \qquad v_y = \frac{d}{dt}(-bt^2) = -2bt$$

and thus

$$\mathbf{v}(t) = \mathbf{i}a - \mathbf{j}2bt$$

Hence, at $t = 0$ we have $\mathbf{v}(0) = \mathbf{i}a = 2.0\mathbf{i}$ m/s, and correspondingly at $t = 1$ second we find that

$$\mathbf{v}(1\text{ s}) = \mathbf{i}a - 2\mathbf{j}b = (2.0\mathbf{i} - 4.0\mathbf{j})\text{ m/s}$$

The speed v of the particle is $(v_x^2 + v_y^2)^{1/2}$ according to (3-8). Hence, at $t = 0$ its speed is 2.0 m/s and at $t = 1$ second it is

$$[(2.0\text{ m/s})^2 + (4.0\text{ m/s})^2]^{1/2} = 4.5\text{ m/s}$$

Example 3-5 A particle travels with a uniform speed v_0 around a circle of radius R. What is its velocity?

Solution Suppose that in Figure 3-20a at $t = 0$ the particle is at A. Then, by definition of the term speed, at time t it will be at the point B, which is a distance $v_0 t$ as measured along the circle from A.

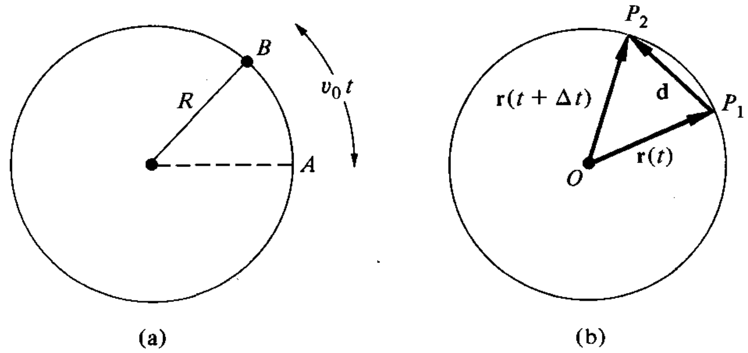


Figure 3-20

To calculate its velocity, let us, as shown in Figure 3-20b, set up an origin at the center of the circle and draw the position vectors $\mathbf{r}(t)$ and $\mathbf{r}(t+\Delta t)$ to the points P_1 and P_2, which specify the position of the particle at the respective times t and $(t+\Delta t)$. As before, the vector from P_1 to P_2 represents the displacement vector $\mathbf{d}$ of the particle. Its average velocity $\bar{\mathbf{v}}$ is given by (3-14); that is,

$$\bar{\mathbf{v}}(t) = \frac{1}{\Delta t}\mathbf{d}$$

and in the limit $\Delta t \to 0$, this quantity approaches the sought-for velocity $\mathbf{v}(t)$. Now in this limit the distance d between P_1 and P_2 approaches the arc length $v_0 \Delta t$ between these same points, and the direction of $\mathbf{d}$ becomes tangent to the circle at P_1. It follows therefore that the velocity $\mathbf{v}$ of this particle has the magnitude v_0 and is directed always tangent to the circle at each point.

Note that even though the speed v_0 of the particle is constant, the velocity is *not*. For as the particle travels around the circle, the *direction* of the velocity is always changing!

3-9 Acceleration

Consider in Figure 3-21a a particle moving along a fixed curve AB. Suppose that at time t, when it is at the point P_1, it has a velocity $\mathbf{v}(t)$ and that at a slightly later time $(t+\Delta t)$, when it is at the point P_2, it has a velocity $\mathbf{v}(t+\Delta t)$. We define the *average acceleration* $\bar{\mathbf{a}}$ of the particle over this time interval to be the change $\Delta\mathbf{v}$ of its velocity per unit time. Thus

$$\bar{\mathbf{a}} = \frac{\Delta \mathbf{v}}{\Delta t} \qquad \text{(3-20)}$$

where $\Delta\mathbf{v}$, as defined in Figure 3-21b, is

$$\Delta\mathbf{v} \equiv \mathbf{v}(t+\Delta t) - \mathbf{v}(t) \qquad \text{(3-21)}$$

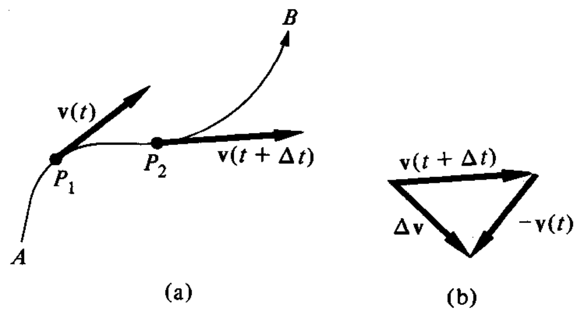


Figure 3-21

It is significant to note that even if the magnitude of the velocity is constant, that is, even if the particle traverses the path AB at constant speed, the velocity difference $\Delta\mathbf{v}$ need *not* vanish. In other words a particle may have an acceleration $\bar{\mathbf{a}}$, even if its speed is constant, provided the direction in which it travels is changing. This important feature will be exemplified in the next section.

The *acceleration* $\mathbf{a}(t)$, or the *instantaneous* acceleration, of a particle is defined to be the limiting value of $\bar{\mathbf{a}}$ as Δt becomes vanishingly small. Thus

$$\begin{aligned}\mathbf{a}(t) &= \lim_{\Delta t \to 0} \bar{\mathbf{a}} \\ &= \frac{d}{dt}\mathbf{v}(t)\end{aligned} \tag{3-22}$$

where the final equality follows by use of (3-20) and the definition of the derivative. Therefore, just as the velocity $\mathbf{v}$ of a particle is the time derivative of its position vector, its acceleration is the time derivative of its velocity. The one-dimensional definitions in (2-6) and (2-9) are manifestly special cases of the more general ones in (3-17) and (3-22), respectively.

For the case that the velocity of the particle has been expressed in terms of basis vectors, it follows from (3-19) that

$$\mathbf{v}(t) = \mathbf{i}v_x + \mathbf{j}v_y$$

Substitution into (3-22) then leads to

$$\mathbf{a}(t) = \mathbf{i}\frac{dv_x}{dt} + \mathbf{j}\frac{dv_y}{dt} \tag{3-23}$$

and therefore the components a_x and a_y of the acceleration in the given system are

$$a_x = \frac{dv_x}{dt} \qquad a_y = \frac{dv_y}{dt} \tag{3-24}$$

Example 3-6 A proton originally traveling at a velocity $\mathbf{v}_0$ enters a region where it is subject to a constant acceleration $\mathbf{a}_0$ at right angles to the initial velocity. Find the position of the particle at any time t.

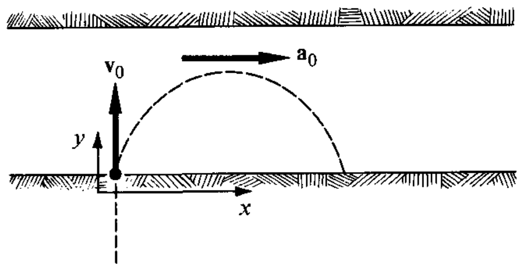


Figure 3-22

Solution As shown in Figure 3-22, let us set up a coordinate system with the origin at the original position of the particle, where it has an initial velocity $\mathbf{v}_0$ directed along the y-axis, and suppose that the constant acceleration $\mathbf{a}_0$ is directed along the x-axis. If $\mathbf{v}(t)$ is the velocity at any time t, then, by (3-22),

$$\frac{d\mathbf{v}}{dt} = \mathbf{i}a_0$$

or, equivalently,

$$\frac{dv_x}{dt} = a_0 \qquad \frac{dv_y}{dt} = 0$$

Integrating, we obtain

$$v_x = a_0 t + c_1 \qquad v_y = c_2$$

where c_1, c_2 are integration constants. Since at $t = 0$ we have $v_x = 0$ and $v_y = v_0$, it follows that $c_1 = 0$ and $c_2 = v_0$. Thus, by the use of (3-19), we obtain

$$\frac{dx}{dt} = a_0 t \qquad \frac{dy}{dt} = v_0$$

Integrating once more, and using the initial condition $x(0) = y(0) = 0$, we find that

$$x(t) = \frac{1}{2}a_0 t^2 \qquad y(t) = v_0 t$$

Therefore the position vector $\mathbf{r}(t)$ is

$$\mathbf{r}(t) = \mathbf{i}\frac{1}{2}a_0 t^2 + \mathbf{j}v_0 t$$

3-10 Centripetal acceleration

It is very important to be aware of the fact that the definition for acceleration in (3-20) through (3-22) involves the difference of two *vectors* and not of two scalars. An immediate consequence of this is that a particle, although moving at a *uniform* speed, may still be accelerating provided the direction of its velocity changes. The purpose of this section is to illustrate the matter by reference to the special but interesting case of a particle moving uniformly in a circular path.

Consider in Figure 3-23 a particle traveling with a uniform speed v_0 counterclockwise in a circular orbit of radius R. As we saw in Example 3-5, the velocity of the particle is at each point of its orbit tangent to the circle. We shall now establish that—as a consequence of this constant change in direction of velocity—as the particle goes around the circle it undergoes a certain acceleration $\mathbf{a}_c$, whose magnitude is

$$a_c = \frac{v_0^2}{R} \tag{3-25}$$

and whose direction is radially inward toward the center of the circle from the instantaneous position of the particle. The term *centripetal acceleration* is reserved for this special type of accelerated motion.

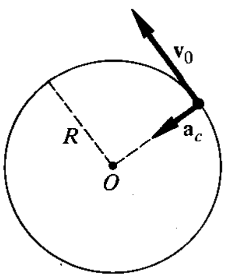


Figure 3-23

To establish the validity of this form for $\mathbf{a}_c$, suppose that in Figure 3-24a the particle is at P_1 at time t and that at the slightly later instant $(t + \Delta t)$ it is at P_2. By hypothesis, the magnitudes of the particle's velocity at both P_1 and P_2 are the same and have the value v_0; that is, both $\mathbf{v}(t)$ and $\mathbf{v}(t + \Delta t)$ have the magnitude v_0 for all Δt. To calculate the acceleration, let us construct the vector difference $\Delta\mathbf{v} = \mathbf{v}(t + \Delta t) - \mathbf{v}(t)$, as shown in Figure 3-24b. Note first that as Δt tends to zero, the two vectors $\mathbf{v}(t + \Delta t)$ and $\mathbf{v}(t)$ tend to become parallel to each other. Therefore, in accordance with the definition in (3-22), as $\Delta t \to 0$, the difference vector $\Delta\mathbf{v}$ (and hence $\mathbf{a}_c$) tends to point from the instantaneous position of the particle toward the center of the circle. To calculate the magnitude of $\mathbf{a}_c$, let us note that for sufficiently small Δt (for

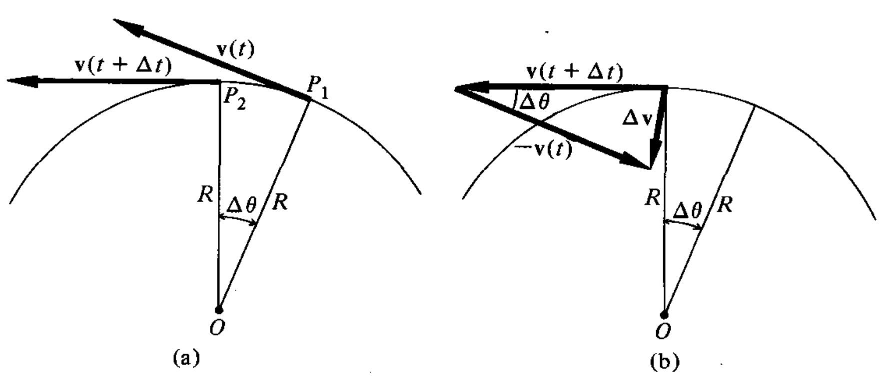


Figure 3-24

which the arc and the chord joining P_1 and P_2 differ by a negligible amount) the triangle OP_1P_2 and that formed by the vectors $\Delta\mathbf{v}$, $\mathbf{v}(t+\Delta t)$, and $-\mathbf{v}(t)$ are similar. For they are both isosceles triangles with vertex angle $\Delta\theta$. Hence, since $|\mathbf{v}(t)| = v_0$, it follows that

$$\frac{\Delta v}{v_0} = \frac{v_0\,\Delta t}{R}$$

since $v_0\,\Delta t$ is the length of the arc (chord) joining P_1 and P_2. Solving for the ratio $\Delta v/\Delta t$ and taking the limit $\Delta t \to 0$, we find by use of (3-22) that the magnitude of this acceleration is indeed v_0^2/R. The validity of (3-25) is thereby established.

The problem of a particle moving nonuniformly on a circular path will be considered in Chapter 5. It will be established there that in addition to the centripetal acceleration the particle may have a *tangential acceleration* as well.

Example 3-7 A bead slides with a uniform speed of 0.5 m/s around a circular ring of radius 20 cm. What is its acceleration?

Solution Substituting the given values, $v_0 = 0.5$ m/s and $R = 0.2$ meter, into (3-25), we obtain

$$a_c = \frac{v_0^2}{R} = \frac{(0.5\text{ m/s})^2}{0.2\text{ m}} = 1.25\text{ m/s}^2$$

The direction of this acceleration is radially inward from the bead toward the center of the circle.

Example 3-8 A child sits at the edge of a merry-go-round of radius 5.0 meters and rotating at 3 revolutions per minute (rpm). What is the child's centripetal acceleration?

Solution In 1 min the child makes three circular orbits and thus travels a distance $3 \times (2\pi R) = 6\pi R$. Hence, since $R = 5.0$ meters, the child's speed is

$$v_0 = \frac{6\pi R}{60\text{ s}} = \frac{6\pi \times 5.0\text{ m}}{60\text{ s}} = 1.6\text{ m/s}$$

and substitution into (3-25) leads to

$$a_c = \frac{v_0^2}{R} = \frac{(1.6\text{ m/s})^2}{5.0\text{ m}} = 0.51\text{ m/s}^2$$

3-11 Motion under constant acceleration

In Section 2-8 we analyzed the vertical motion of a particle under the influence of the uniform gravitational field of say a planet. The purpose of this section is to generalize the results in (2-15) through (2-17) to the more general case in which the particle has a horizontal component of motion at the same time that it is in free fall.

Suppose a particle is initially at a height h above the ground and that it has at this instant a velocity $\mathbf{v}_0$ which makes an angle α with the horizontal. Let us set up a coordinate system with the y-axis vertically upward and with the horizontal x-axis selected in such a way that the initial velocity vector $\mathbf{v}_0$ lies in the x-y plane. Further, assume as shown in Figure 3-25 that the origin of the coordinate system is selected so that initially the particle has the coordinates $(0, h)$.

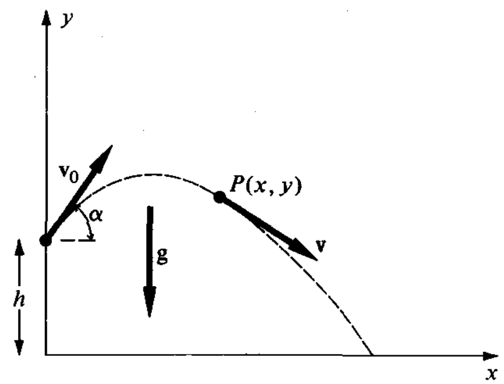


Figure 3-25

In terms of the given coordinate system the acceleration of the particle is $-\mathbf{j}g$ with g given in (2-14). Substitution into (3-24) shows then that the two-dimensional analogue of (2-15) is

$$\frac{dv_x}{dt} = 0 \qquad \frac{dv_y}{dt} = -g$$

with $g = 9.8\ \mathrm{m/s^2}$.

Proceeding now as in Example 3-6, we obtain the trajectory of the particle—the dashed curve in Figure 3-25—by integrating these relations. This time, the initial conditions are $v_x(0) = v_0 \cos\alpha$, $x(0) = 0$, $v_y(0) = v_0 \sin\alpha$, $y(0) = h$, and the final result is

$$\begin{aligned} x(t) &= tv_0 \cos\alpha \\ y(t) &= h - \frac{1}{2}gt^2 + tv_0 \sin\alpha \end{aligned} \tag{3-26}$$

Differentiating, we obtain the associated velocity formulas

$$\begin{aligned} v_x &= v_0 \cos\alpha \\ v_y &= -gt + v_0 \sin\alpha \end{aligned} \tag{3-27}$$

whose time derivative yields $a_x = 0$, $a_y = -g$, as it should. An examination of these formulas shows that motion in a uniform gravitational field can be thought of as consisting of a horizontal motion at the uniform velocity $v_x = v_0 \cos\alpha$ and of a uniformly accelerated motion along the vertical. This feature is illustrated by the multiflash photograph in Figure 3-26, which shows two golf balls, one of which falls freely under gravity while the other is given

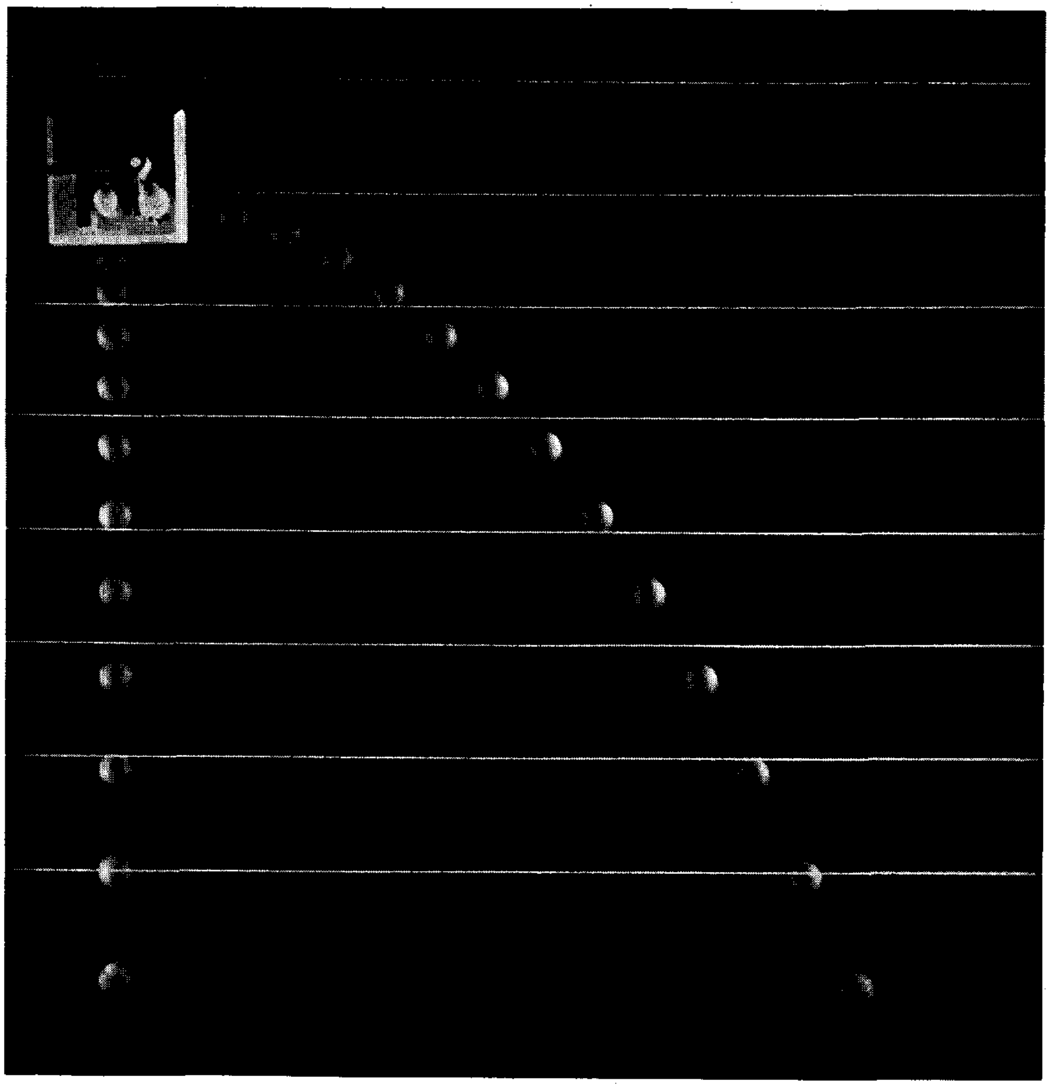


Figure 3-26 *A flash photograph of two golf balls released simultaneously, one with a horizontal velocity of 2.0 m/s and the second from rest. The light flashes are triggered at intervals of 1/30 of a second. The horizontal white lines are a series of parallel strings behind the golf balls and 15 cm apart. (Reprinted from College Physics, with permission of D.C. Heath and Company.)*

an initial horizontal component of velocity. The vertical displacement of the two are precisely the same at all times.

For the special case, $\alpha = 90°$, the initial velocity $\mathbf{v}_0$ is directed vertically upward, and the entire motion should take place along the y-axis. Indeed, using the facts $\cos 90° = 0$, $\sin 90° = 1$, it is easy to confirm that (3-26) and (3-27) reduce to our previous results in (2-16) and (2-17).

An alternate derivation of (3-26) involving the use of the Galilean transformation will be found in Problem 39.

The curve in space along which the particle actually moves may be obtained by eliminating t between the two relations in (3-26). Solving the first for t and

substituting into the second, we find

$$y = h + x \tan \alpha - \frac{g}{2(v_0 \cos \alpha)^2} x^2 \tag{3-28}$$

and Figure 3-27 shows a plot of this curve for the special case $h = 0$, for which the body is projected up from the ground. In constructing this graph, we have used the fact that $\sin 2\alpha = 2 \sin \alpha \cos \alpha$ and that its maximum occurs at the point $[v_0^2(\sin 2\alpha)/2g,\ v_0^2(\sin^2 \alpha)/2g]$, at which $dy/dx = 0$.

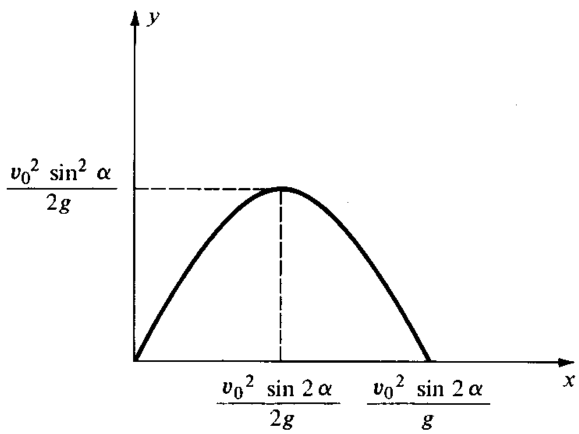


Figure 3-27

For the special case shown in Figure 3-27, in which the projectile is projected from the surface of the earth, an important parameter is the *range R* of the projectile. The range is defined to be the horizontal distance between the point on the ground where the projectile goes up and the point where it again strikes the ground. Reference to the figure shows that the range R is

$$R = \frac{v_0^2}{g} \sin 2\alpha \tag{3-29}$$

and thus for any given value for $\mathbf{v}_0$ there is a unique range. Since the maximum value of $\sin 2\alpha$ is unity and this occurs when $2\alpha = 90°$ or $\alpha = 45°$, it follows that, for a fixed value for v_0, the maximum range occurs at an elevation angle of $\alpha = 45°$. Furthermore, the maximum range R_{max} for a given velocity v_0 is

$$R_{max} = \frac{v_0^2}{g} \tag{3-30}$$

Example 3-9 A golf ball is launched from the ground at an angle of 60° with respect to the horizontal and with a velocity of 50 m/s.

(a) What is its velocity at the height of its trajectory?
(b) What is its maximum height above the ground?
(c) What is its range?

Solution Substitution of the given data ($\alpha = 60°$, $v_0 = 50$ m/s, $h = 0$, $\cos 60° = 0.50$, $\sin 60° = 0.87$) leads to

$$x(t) = 25t$$
$$y(t) = -4.9t^2 + 43t$$

where x and y are in meters and t is in seconds.

(a) At the highest point of its path, the velocity of the golf ball is directed along the horizontal. Differentiating $x(t) = 25t$, we find that

$$v_x = \frac{dx}{dt} = 25 \text{ m/s}$$

(b) According to Figure 3-27, the maximum height y_{max} above the ground is

$$y_{max} = \frac{v_0^2 \sin^2 \alpha}{2g} = \frac{(50 \text{ m/s})^2 \times (0.87)^2}{2 \times 9.8 \text{ m/s}^2} = 96 \text{ m}$$

(c) The substitution of the given data into (3-29) yields

$$R = \frac{v_0^2 \sin 2\alpha}{g} = \frac{(50 \text{ m/s})^2 \times \sin 120°}{9.8 \text{ m/s}^2} = 220 \text{ m}$$

since $\sin 120° = \cos 30° = 0.87$.

Example 3-10 Suppose that a bomb is released from a bomber flying at an elevation of 4.0 km at a horizontal speed of 10^3 km/hr. How long after being dropped does the bomb strike? What is its velocity at impact? (See Figure 3-28.)

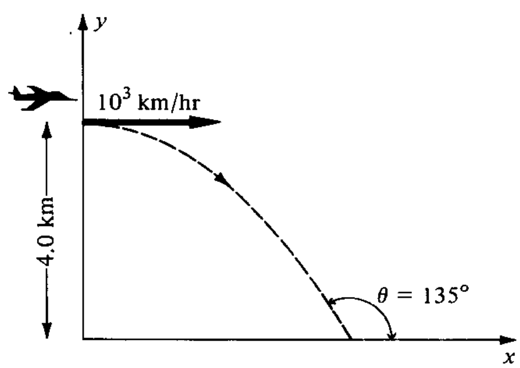


Figure 3-28

Solution At the instant of release, the bomb has the same velocity as does the plane. Hence the appropriate parameter values are $\alpha = 0$, $h = 4.0 \times 10^3$ meters, and $v_0 = 1.0 \times 10^3$ km/hr $= 280$ m/s.

If τ is the time when the bomb strikes the ground, we find by the use of (3-26) that

$$y(\tau) = 0 = h - \frac{1}{2}g\tau^2$$

since $\sin \alpha = 0$ in this case. Hence

$$\tau = \left[\frac{2h}{g}\right]^{1/2} = \left[\frac{2 \times 4.0 \times 10^3 \text{ m}}{9.8 \text{ m/s}^2}\right]^{1/2}$$
$$= 29 \text{ s}$$

At this instant $t = \tau$ its horizontal velocity v_x is still $v_0 = 280$ m/s and, according to the second equation of (3-27),

$$v_y(\tau) = -g\tau = -(9.8\ \text{m/s}^2) \times 29\ \text{s}$$
$$= -2.8 \times 10^2\ \text{m/s}$$

Hence the speed v at impact is

$$v = [v_x^2 + v_y^2]^{1/2} = [(280\ \text{m/s})^2 + (280\ \text{m/s})^2]^{1/2}$$
$$= 400\ \text{m/s}$$

and its direction of travel θ is determined by

$$\tan\theta = \frac{v_y}{v_x} = \frac{-280\ \text{m/s}}{280\ \text{m/s}} = -1$$

and thus $\theta = 135°$.

†3-12 The Lorentz transformation

Let us reexamine the analysis in Section 2-9 of the motion of a particle as seen by two observers in relative motion. This time, however, let us treat the more general case, in which the motion of the particle is not confined to a straight line but instead is free to move in a plane.

To this end, consider in Figure 3-29 two observers O and O', who are at rest relative to their respective coordinate systems S and S', and who observe the motion of a particle P along some planar trajectory AB. Assuming that S and S' are in relative motion at the constant velocity $\mathbf{u}$, let us select parallel sets of coordinate axes so that the trajectory AB lies in the common x-y and x'-y' planes and so that $\mathbf{u}$ lies along the positive senses of the parallel x- and x'-axes. For reasons to be clarified below, suppose that there are two identical clocks C and C' at rest in S and S', respectively, and that, at the instant when the origins of S' and S coincide, both clocks are set to read "zero time." That is, if t and t' represent the respective readings of the two clocks at any subsequent instant, then at $t = t' = 0$, the origins of S and S' coincide. Since

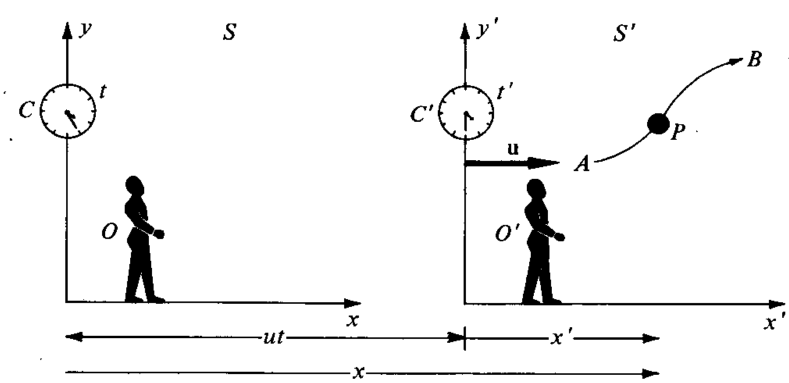


Figure 3-29

the clocks are identical, we expect that the readings of C and C' will coincide at all times. However, to focus attention on the fact that this is indeed an assumption, let us continue this distinction between t and t'.

Consider now the description of the motion of P according to these two observers. The observer O, who is at rest in S, describes the motion of the particle in terms of the coordinates $x = x(t)$, $y = y(t)$ in S and with the time as measured on C. Similarly, the other observer describes this same motion in terms of the coordinates $x' = x'(t')$, $y' = y'(t')$ as measured relative to his axes and according to time t' as measured on his clock. Repeating the arguments of Section 2-9 and making use of "common sense" ideas of space and time, we find that these coordinates and times as seen by these two observers are related by

$$\begin{aligned} x' &= x - ut \\ y' &= y \\ t' &= t \end{aligned} \tag{3-31}$$

The first two of these relations are an obvious generalization of (2-19). They are a direct consequence of the geometrical relations in Figure 3-29 and constitute the two-dimensional form of the *Galilean transformation.* The last equation of (3-31) has been added to emphasize the assumption that two clocks in relative motion run at the same rate.

Prior to the year 1905 it was generally believed that this Galilean transformation in (3-31) was the only correct way to relate the descriptions for the motion of the particle as seen by two observers in relative motion. Indeed, only "self-evident" geometrical notions were required in the derivation above. Moreover, experiments involving speeds small compared to the speed of light, $c\,(\cong 3.0 \times 10^8$ m/s), have shown that the relations in (3-31) are fully in accord with our expectations. When experiments were extended to speeds of the order of c, however, it was discovered that the relations in (3-31) were completely at variance with the observed facts. This surprising and unexpected feature was cleared up in a decisive way in 1905 by a certain clerk in the Swiss Patent Office named Albert Einstein. When only 26 years old he proposed his "Special Theory of Relativity,"[1] and in one bold epoch-making stroke he showed how we must modify our space-time concepts so that they would be again in harmony with experimental realities. One of the very important consequences of relativity theory is that the Galilean transformation is *not* generally valid. Instead, the correct transformation is one that had been proposed earlier by H. A. Lorentz (1853–1928) and is now known as the *Lorentz transformation.* A very striking feature of this transformation is that the clocks C and C' in Figure 3-29 will not remain in step except if $u = 0$.

To define the Lorentz transformation, suppose in Figure 3-29 that the

[1]A. Einstein, *Annalen Physik,* vol. 17, p. 891, 1905.

relative velocity between S and S' is the constant u. Define the parameter γ by

$$\gamma = \frac{1}{[1 - u^2/c^2]^{1/2}} \tag{3-32}$$

with c the speed of light. Note that u can assume all values in the range $-c < u < c$, where negative values for u correspond in Figure 3-29 to S' traveling along the negative x-axis. The fact that *no* two systems can travel at a relative velocity larger than c, and that thus imaginary values of γ do not occur, is an experimental fact which is fully consistent with the theory of relativity. In terms of γ, and for the situation in Figure 3-29 the Lorentz transformation is

$$\begin{aligned} x' &= \gamma(x - ut) \\ y' &= y \\ t' &= \gamma\left(t - \frac{u}{c^2}x\right) \end{aligned} \tag{3-33}$$

and, just as for the Galilean transformation, given the coordinates x and y in S of the particle at any time t, we may calculate the coordinates (x', y') in S' of the particle at any time t'. Note that, in general, the times t' and t in (3-33) are *not* the same, and indeed the relation between t and t' generally depends on the position of the particle. It is left as an exercise to confirm that for $u \ll c$, and for particle velocities small compared to c, the relations in (3-33) reduce to those in (3-31). Hence the Galilean transformation is the low-velocity limit of the Lorentz transformation.

In closing this very brief discussion of the Lorentz transformation, let us derive by its usage the very interesting *addition law for velocities*. To this end, suppose that the particle in Figure 3-29 is undergoing one-dimensional motion along the parallel x- and x'-axes. If in a time interval Δt, as seen in S, it undergoes a displacement Δx, then the associated displacement $\Delta x'$ and time interval $\Delta t'$ in S' will be related to Δx and Δt in accordance with (3-33) by

$$\Delta x' = \gamma(\Delta x - u\Delta t)$$

$$\Delta t' = \gamma\left(\Delta t - \frac{u}{c^2}\Delta x\right)$$

Dividing the second of these into the first, we obtain

$$\frac{\Delta x'}{\Delta t'} = \frac{\dfrac{\Delta x}{\Delta t} - u}{1 - \dfrac{u}{c^2}\dfrac{\Delta x}{\Delta t}} \tag{3-34}$$

where on the right-hand side both numerator and denominator have been divided by Δt. Now the velocity v of the particle as seen by O is

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$

while the corresponding velocity v' as seen by O' is dx'/dt'. Hence, if we take the limit $\Delta t \to 0$, (3-34) becomes

$$v' = \frac{v - u}{1-(vu/c^2)} \tag{3-35}$$

and this is the velocity addition theorem. Note that for the special case in which $u \ll c$, and $v \ll c$, this reduces to the previous Galilean transformation result in (2-20).

One of the very interesting implications of (3-35) is that if a particle (photon) travels at the speed of light in S, then it also travels at this same speed c in S' regardless of the relative velocity u. For, setting $v = c$ in (3-35) we find directly that

$$v' = \frac{c - u}{1-(cu/c^2)} = c$$

It is this very unexpected fact—namely, that the velocity of light is the same regardless of the relative velocity between the source of light and its observer—that has played a key role in the development of the theory of relativity. Indeed, this feature was one of the two basic hypotheses which Einstein used as a basis for his theory.

3-13 Summary of important formulas

Given a vector **A**, we define its projections A_x, A_y on the axes of a coordinate system by

$$A_x = A \cos \theta \qquad A_y = A \sin \theta \tag{3-7}$$

where A is the magnitude of **A** and θ is the angle between **A** and the positive sense of the x-axis. Conversely, given the components B_x, B_y of a vector, we may reconstruct the vector by the formulas

$$B = \sqrt{B_x^2 + B_y^2} \tag{3-8}$$

$$\theta = \tan^{-1} \frac{B_y}{B_x} \tag{3-9}$$

If a particle undergoes a displacement $\Delta \mathbf{r} = \mathbf{r}(t + \Delta t) - \mathbf{r}(t)$ in a time interval Δt, then its *average* velocity $\bar{\mathbf{v}}$ over the interval is

$$\bar{\mathbf{v}} = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} \tag{3-15}$$

and the limit of $\bar{\mathbf{v}}$ as $\Delta t \to 0$ is the *instantaneous velocity* **v**:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} \tag{3-17}$$

Similarly, if $\Delta\mathbf{v}$ is the change in the velocity of a particle during a time interval Δt, then its *average acceleration* $\bar{\mathbf{a}}$ over the interval is

$$\bar{\mathbf{a}} = \frac{\Delta\mathbf{v}}{\Delta t} \tag{3-20}$$

and its acceleration $\mathbf{a}$ is

$$\mathbf{a} = \lim_{\Delta t \to 0} \bar{\mathbf{a}} = \frac{d\mathbf{v}}{dt} \tag{3-22}$$

If a particle travels at a uniform speed v_0 around a circle of radius R, it undergoes a centripetal acceleration $\mathbf{a}_c$, which is directed radially inward from the instantaneous position of the particle and has the magnitude

$$a_c = \frac{v_0^{\,2}}{R} \tag{3-25}$$

The trajectory of a particle launched as shown in Figure 3-25 and subsequently under uniform acceleration $a_x = 0$, $a_y = -g$ is

$$x(t) = tv_0 \cos\alpha$$

$$y(t) = h - \frac{1}{2}gt^2 + tv_0 \sin\alpha \tag{3-26}$$

QUESTIONS

1. Define or describe briefly what is meant by the following terms: (a) scalar (b) vector; (c) unit vector; (d) basis vectors; (e) components of a vector; and (f) vector function.
2. Give two examples of scalars and two examples of vectors.
3. What must be the relationship between two vectors if their vector sum or resultant is the null vector?
4. What must be the relationship between three vectors if their vector sum vanishes? What must be this relationship if there are n vectors?
5. Suppose A_x, A_y are the components of a vector in a certain Cartesian coordinate system. What are the components A_x', A_y' of $\mathbf{A}$ in an x', y' system that is related to the original system by the transformation $x' = -x$; $y' = y$?
6. A particle moves with constant speed in a straight line. Does it accelerate? Explain.
7. A bead is forced to slide with uniform speed on a wire in the form of an ellipse. Does it accelerate? Explain.
8. Is it possible for the speed of an automobile to change if the automobile is not accelerating? Can an accelerating particle have zero velocity?
9. Consider the particle moving along the trajectory AB in Figure 3-21. Why must the vectors $\mathbf{v}(t)$ and $\mathbf{v}(t+\Delta t)$ be tangent to the path? Is it possible for $\mathbf{v}$ not to be tangent to such a curve? Explain.
10. An automobile is driven at a uniform speed around a bend in a road, which is an arc of a circle. Explain why the automobile accelerates and describe the direction of this acceleration.
11. A horizontally pitched baseball is struck by a batter and flies backward along its incident trajectory. What is the direction of its acceleration during this process? If it were popped

vertically upward with its incident speed, what would now be the direction of its acceleration?

12. Give the direction of the acceleration of an elevator if (a) it starts to rise; (b) it stops after having risen; (c) it starts to descend; and (d) it stops after having descended.
13. X rays are usually produced commercially by causing rapidly moving electrons to collide with an electrode and thereby slowing down. What is the direction of their acceleration in such a collision?
14. A particle slides around a circular track. Is it ever possible for its acceleration to be purely tangential? Under what circumstances will its acceleration be radial?
15. A boy sits in a train, which is moving uniformly along a straight horizontal track at a speed of 90 km/hr, and amuses himself by throwing a ball vertically upward and catching it. What is the trajectory of the ball as seen by an observer standing on the ground?
16. A boy standing at the outer edge of a rotating merry-go-round throws a ball vertically upward. He finds that it does not appear to rise and to descend vertically as did the boy in Question 15. Explain.
17. Consider again the situation in Question 16. Explain why to a stationary observer, *not* on the merry-go-round, the trajectory of the ball is a parabola.
18. An automobile, which is traveling parallel to a uniformly moving train is observed by two observers O_1 and O_2, with O_1 sitting on the train and O_2 sitting on a porch nearby. The automobile suddenly undergoes an acceleration $\mathbf{a}_0$. (a) What acceleration of the auto is observed by O_1? (b) What acceleration of the auto is observed by O_2? (c) What acceleration of O_1 and of O_2 is observed by the driver of the auto?
19. Two observers O_1 and O_2 measure the velocity of a particle and obtain the same results. Based on these data, what can you say about the relative motion of O_1 and O_2?
20. Suppose that now the two observers of Question 19 measure different values for the velocity, but the same value for the acceleration. What can you say now about the relative motion between O_1 and O_2? What could you say if they also measured different values for the acceleration?

PROBLEMS

1. Three vectors **A**, **B**, and **C** correspond, respectively, to the following velocities: 4 km/hr, due north; 8 km/hr, due east; and 6 km/hr, 45° south of east. Select a convenient scale (such as 1 cm corresponds to 1 km/hr).
   (a) On paper draw arrows representing the vectors **A**, **B**, and **C**.
   (b) On the same paper draw arrows for the vectors $-\mathbf{A}, 0.75\mathbf{B}, -2\mathbf{C}$.
2. For the vectors **A**, **B**, **C**, as defined in Problem 1 draw an arrow for the vectors (a) $\mathbf{A}+\mathbf{B}$; (b) $2\mathbf{A}-\mathbf{B}$; (c) $\mathbf{A}+\mathbf{B}+\mathbf{C}$.
3. Consider again the vectors **A**, **B**, **C**, of Problem 1. (a) Draw the component of **C** along **B** and measure its length. Express your answer in kilometers per hour. (b) Draw the component of **B** along **C**, measure its length, and express it in kilometers per hour. Should this be the same as your answer to (a)? Explain.
4. Three displacement vectors $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_3$ have the respective magnitudes of 1 meter, 2 meters, and 3 meters and are oriented so that the angle between $\mathbf{d}_1$ and $\mathbf{d}_3$ is 90° and $\mathbf{d}_2$

lies between them in their plane and makes an angle of 30° with $\mathbf{d}_1$.

(a) Select a scale and draw these vectors.

(b) Draw the component of $\mathbf{d}_3$ along $\mathbf{d}_2$ and express its length in meters.

(c) Through what distance will a particle be displaced if it undergoes the displacement $(\mathbf{d}_1 + \mathbf{d}_2)$?

5. Two acceleration vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are at right angles to each other and have the magnitudes $10\ \mathrm{m/s^2}$ and $15\ \mathrm{m/s^2}$, respectively.
   (a) Draw these two vectors, to some scale, on paper.
   (b) Draw the vector $(\mathbf{a}_1 + \mathbf{a}_2)$ and express its length in meters per second squared.
   (c) Repeat (b) for the vectors $(\mathbf{a}_1 - \mathbf{a}_2)$ and $(2\mathbf{a}_1 + 2\mathbf{a}_2)$.
6. An automobile is driven at 100 km/hr due north for 1 hour and then for an additional 45 min at the same speed in a direction 45° east of north. What is the displacement vector of the automobile during the total time interval? What is its average speed?
7. An airplane flies due east for 1 hour at a speed of 500 km/hr and then it flies at this same speed in a certain direction so that after an additional 1 hour of flying it is at a distance of 800 km from its starting point. If the final location of the airplane is south of its original position, find the direction in which it was flown during the second hour.
8. Two acceleration vectors **A** and **B** have the respective magnitudes $3\ \mathrm{m/s^2}$ and $4\ \mathrm{m/s^2}$ and an angle of 60° between them. What is the magnitude of the vector $(\mathbf{A}+\mathbf{B})$? What is the angle between $(\mathbf{A}+\mathbf{B})$ and **A** and between $(\mathbf{A}+\mathbf{B})$ and **B**?
9. Repeat Problem 8, but this time for the vector $(\mathbf{A}-\mathbf{B})$.
10. (a) Prove that the magnitude $|\mathbf{A}+\mathbf{B}|$ of the two vectors **A** and **B** is

$$|\mathbf{A}+\mathbf{B}| = [A^2 + B^2 + 2AB\cos\theta]^{1/2}$$

where θ is the angle between the two vectors (see Figure 3-30).

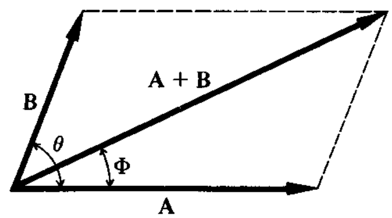


Figure 3-30

(b) Prove also that the angle Φ between the vectors $(\mathbf{A}+\mathbf{B})$ and **A** satisfies

$$\tan\Phi = \frac{B\sin\theta}{A + B\cos\theta}$$

11. In a certain coordinate system two vectors **A** and **B** have the form

$$\mathbf{A} = 3\mathbf{i} - 4\mathbf{j} \qquad \mathbf{B} = 2\mathbf{i} + 3\mathbf{j}$$

(a) What are the components of the vector $(\mathbf{A}+\mathbf{B})$ in this system? (b) What is the component of $(\mathbf{A}+2\mathbf{B})$ along the x-axis? Along the y-axis? (c) What is the component of $-\mathbf{A}$ along the x-axis? Along the negative y-axis?

12. Consider again the two vectors **A** and **B** of Problem 11. (a) What is the magnitude of **A**? (b) What is the magnitude of **B**? (c) Write down formulas for unit vectors parallel to **A** and to **B**.
13. Let B_x, B_y be the components of a vector **B** along the axes of a certain coordinate system and B_x', B_y' the corresponding components along the axes of an x'-y' system that is rotated with respect to the original one by an angle α (see Figure 3-31). Prove that

$$B_x' = B_x \cos\alpha + B_y \sin\alpha$$
$$B_y' = -B_x \sin\alpha + B_y \cos\alpha$$

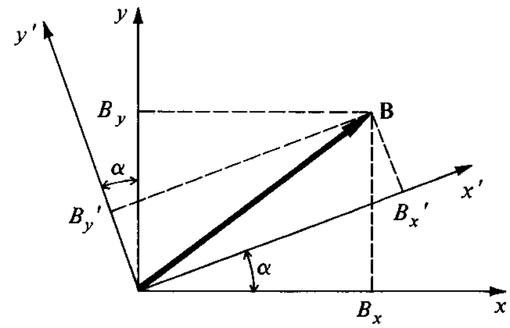


Figure 3-31

14. Show that if B_x, B_y and B_x', B_y' are defined as in Problem 13, then

$$(B_x')^2 + (B_y')^2 = B_x^2 + B_y^2$$

or, in other words, that the magnitude of a vector is invariant under rotations of the coordinate axes. Explain in physical terms why this must be so.

15. A vector **F** has in a certain coordinate system the form

$$\mathbf{F} = 3\mathbf{i} + 2\mathbf{j}$$

(a) What is $|\mathbf{F}|$? (b) What is the angle between **F** and the positive x-axis? (c) Write down a formula for a unit vector parallel to **F**.

16. In the same coordinate system as in Problem 15 a vector **G** has the form $\mathbf{G} = 2\mathbf{i} - \alpha\mathbf{j}$ with α some parameter.
    (a) Determine the value of α such that **G** is parallel to **F**. (*Hint*: Determine the angle between **G** and the x-axis.)
    (b) Determine the value of α such that **G** is perpendicular to **F**.
17. Prove that if two vectors **A** and **B** have the respective components A_x, A_y and B_x, B_y, then (a) if $A_xB_y = A_yB_x$, then **A** and **B** are parallel and (b) if $A_xB_x = -A_yB_y$, then **A** and **B** are perpendicular.
18. In a certain coordinate system, the position $\mathbf{r}(t)$ (in meters) of a particle is given by $\mathbf{r}(t) = \mathbf{i}t^2 - 2\mathbf{j}t^3$, with t in seconds.
    (a) What is its velocity **v** at any time t?
    (b) What is its speed at $t = 1$ second?
    (c) Calculate the magnitude of its acceleration at any time t.
19. The velocity of a particle in a certain coordinate system is $\mathbf{v}(t) = 3\mathbf{i}t - \mathbf{j}t^2$, with t in seconds and $|\mathbf{v}|$ in meters per second. If the particle is originally at the origin, calculate its position at any time t and check your answer.
20. A particle is initially at the origin of a certain coordinate system and has (in meters per second) the initial velocity $\mathbf{v}_0 = 2\mathbf{i} - \mathbf{j}$. If it suffers a constant acceleration $\mathbf{a} = 3\mathbf{i}\ \mathrm{m/s^2}$, calculate its velocity and position at any subsequent time t.
21. A muon is traveling at a velocity of 3.0×10^5 cm/s when it enters the region between two capacitor plates, where it experiences a constant acceleration of $6.0 \times 10^{10}\ \mathrm{cm/s^2}$ directed as shown in Figure 3-32. Suppose that the direction of the incident velocity makes an angle of 150° with the direction of the acceleration.
    (a) Set up a coordinate system and express the acceleration $d\mathbf{v}/dt$ of the particle in your system.
    (b) Calculate $\mathbf{v}(t)$ for the muon while it is between the capacitor plates.
    (c) Find $\mathbf{r}(t)$ for the muon.

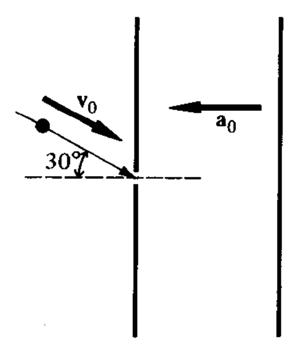


Figure 3-32

22. Calculate the acceleration of a point on the rim of a wheel of radius 30 cm which rotates about its axis at 10 rps (revolutions per second).
23. By virtue of the earth's rotation about its axis, a person located on the earth at a latitude λ moves on the rim of a circle of radius $R \cos \lambda$ where $R (\cong 6.4 \times 10^3 \text{ km})$ is the radius of the earth (see Figure 3-33).

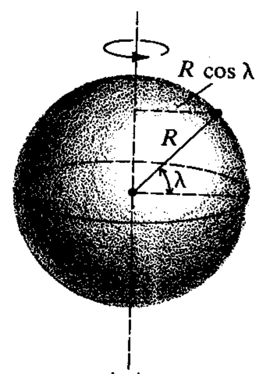


Figure 3-33

Calculate his centripetal acceleration in terms of λ and the period of rotation of the earth P.
24. During the flight of Apollo 11, while Armstrong and Aldrin descended to the lunar surface in the LEM, Collins continued to orbit the moon in a nearly circular orbit about 90 km above its surface. Assuming that it took him 1 hour to completely circle the moon once, and that he was at a distance of 1.7×10^3 km from the lunar center, calculate his centripetal acceleration.
25. A child sits at a distance of 4 meters from the axis of a merry-go-round, which makes one complete turn in 10 seconds. If the child walks radially inward a distance of 2 meters, by what amount has its centripetal acceleration been decreased?
26. A particle is thrown up from the ground with an initial velocity of 10 m/s in a direction making 45° with the horizontal. Calculate:
    (a) The particle's maximum height above the ground.
    (b) The time required to reach its greatest height.
    (c) Its velocity at this instant.
27. For the situation described in Problem 26, calculate (a) the particle's range and (b) the time interval during which it is in flight.
28. A small sphere rolls at a certain velocity $\mathbf{v}_0$ toward the edge of a table 1.2 meters in height. Calculate the value of v_0 if it strikes the ground at a point 2 meters from the foot of the table. For how long a time is the sphere in free fall?
29. A projectile is fired upward at an angle α with respect to the horizontal, with a fixed velocity $\mathbf{v}_0$. Show that for elevation angles that exceed or fall short of $\alpha = 45°$ by equal amounts, the range is the same. (*Note*: This result was first obtained by Galileo.)
30. What must be the minimum speed of a football at kickoff so that it travels 60 yd (about 55 meters) to the opponent's goal line? Where would this football have landed if it had the same initial velocity but were directed at an angle of 60° with the horizontal?
31. A golf ball is struck in such a way that its initial velocity makes an angle of 45° with the horizontal. If it is observed to land on the green 180 meters away, calculate its initial speed.

*32. Suppose that the maximum horizontal range of a certain cannon for a fixed muzzle velocity is R_0.
    (a) Show that the muzzle velocity v_0 associated with this cannon is

$$v_0 = \sqrt{gR_0}$$

    (b) Suppose that this cannon is at the foot of a hill of elevation angle β and is fired at an angle α with respect to the hill (see Fig-

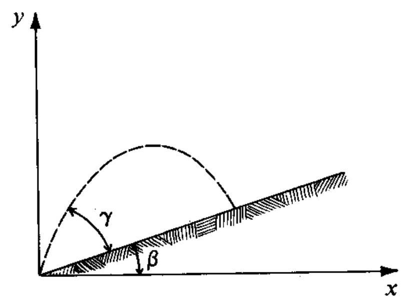


Figure 3-34

ure 3-34). Show that the trajectory of the projectile may be expressed in the given coordinate system in the form

$$y = x \tan(\beta + \gamma) - \frac{1}{2}\frac{x^2}{R_0 \cos^2(\beta + \gamma)}$$

***33.** In terms of the parameters β, γ, and R_0 in Problem 32, calculate the range R of the projectile as measured along the hill. Find the value of the angle γ for which this range is a maximum.

34. A cannon fires a shell up a hill of elevation angle $\beta = 30°$ (see Figure 3-34). If the muzzle velocity v_0 of the projectile is 700 m/s and the angle the cannon makes with the hill is 15°, calculate **(a)** the range of the projectile along the hill and **(b)** the time that the shell is in flight.

35. A baseball is thrown upward from a height of 2 meters above the ground with an initial velocity of 20 m/s in a direction making an angle of 45° with the horizontal.

(a) What is the maximum height reached by the ball?

(b) How far away (measured along the ground from the place where it was originally thrown) is the baseball when it strikes the ground?

(c) What is its velocity at impact?

†36. A person stands 1.5 meters above the ground on the rear end of a train that is traveling at a uniform velocity of 90 km/hr. He ejects an object from the floor of the train, in the direction from which the train has just come, with a speed (relative to the train) of 10 m/s and at an angle of 37° with respect to the horizontal.

(a) What is the magnitude and direction of the initial velocity of the object as seen by an observer fixed to the earth?

(b) From the viewpoint of a fixed observer how far and in what direction will the object have traveled before striking the ground?

(c) How long will it be in flight?

†37. The position vector $\mathbf{r}(t)$ for a particle in coordinate system S in Figure 3-29 is

$$\mathbf{r}(t) = \mathbf{i}t^2 + 3\mathbf{j}t^3$$

with $\mathbf{r}$ in meters and t in seconds. Let us assume the validity of the Galilean transformation.

(a) What is the position vector $\mathbf{r}'$ as seen by an observer with respect to whom S travels at the uniform velocity $\mathbf{u} = 2\mathbf{i}$ m/s? Assume that at $t = 0$; $\mathbf{r} = \mathbf{r}' = 0$.

(b) Repeat (a) for the case $\mathbf{u} = -3\mathbf{i}$ m/s.

†38. A boy sitting on a train throws a ball vertically upward so that its position above his hand at any time t is

$$y = -\frac{1}{2}gt^2 + v_0 t$$

where $g = 9.8\ \text{m/s}^2$ is the acceleration of gravity and $v_0 = 5.0$ m/s is the initial velocity. If the train goes at a speed of 150 km/hr, what is the position vector for the ball as seen by an observer stationary with respect to the ground? Assume the validity of (3-31) and Figure 3-29.

†39. Suppose, in Figure 3-29, that the trajectory of the particle as seen in

S' is

$$y' = -\frac{1}{2}gt^2 + tv_0 \sin\alpha - h$$

$$x' = 0$$

(a) What is the observed trajectory in S, assuming the validity of (3-31)?

(b) Show that for $u = v_0 \cos\alpha$ your answer to (a) is precisely the same as (3-26). Explain this result.

†40. Suppose that the observer in S in Figure 3-29 throws an object vertically upward so that its trajectory is $x = 0$; $y = v_0t - \frac{1}{2}gt^2$, with v_0 and g the usual constants. Show by use of the Lorentz transformation that the trajectory in S' is

$$x' = -ut'$$

$$y' = \frac{-1}{2\gamma^2}gt'^2 + \frac{v_0}{\gamma}t'$$

†41. In Figure 3-29, suppose that the particle moves along the trajectory $x = tv_0 \cos\theta$; $y = tv_0 \sin\theta$, with θ a fixed constant. Calculate, by use of (3-33), the position $x'(t')$, $y'(t')$ of this same particle as seen by the observer in S'.

†42. Two spaceships take off in opposite directions from a distant planet and eventually achieve their cruising speeds of $0.8c$. That is, an observer on the planet observes that each ship travels relative to him at a speed of $0.8c$, but in opposite directions. By use of the velocity addition theorem show that the relative velocity between the spaceships is $(1.6/1.64)c$, and compare this with the value predicted by the Galilean transformation.

4 The laws of Newton

A body in motion or at rest must be brought into that state of motion or rest by the action of another body which in turn is brought into its state of motion or rest by a third body, and so on ad infinitum.

BARUCH SPINOZA

4-1 The laws of Newton

We now turn from our study of kinematics and, in this and the following chapters on mechanics, focus attention on the field of *dynamics.* In addition to describing motion, here we shall also be concerned with the "influences" or *forces,* that can alter motion.

Underlying the study of dynamics are Newton's three laws of motion, and the purpose of this chapter is to describe these laws. Initially, we shall formulate these Newtonian laws of mechanics only for a single particle. After obtaining some familiarity with this case, we then apply them in Chapter 11 to the individual members of collections of particles and thus establish that, with some reinterpretation, these single-particle laws are also applicable to macroscopic bodies. Anticipating this more general formulation, we shall, in the interim, freely apply the single-particle dynamical laws to this more general class of phenomena.

4-2 Force—qualitative discussion

Newton's laws of motion involve the three physical quantities of *force, mass,* and *acceleration.* Since an operational definition for acceleration is already at hand we now require correspondingly precise definitions for the concepts of force and mass. By way of introduction to such operational definitions, this section is devoted to a qualitative discussion of force.

Qualitatively speaking, a force is said to act on a body if an agent of some type exerts a push or a pull on it. Typically, a force will cause either a deformation of the body or an alteration in its state of motion or both. Consider, for example, in Figure 4-1a an unstretched spring with one end attached to a vertical wall. If the other end of the spring is now pulled so that it is stretched, as shown in Figure 4-1b, then we say that a force has been exerted on it. Similarly, if a child pushes on a wagon, thereby causing it to accelerate, the child exerts a force on it. Although in many cases when an agent exerts a force on an object he causes a deformation or an acceleration of some type, this need not always be the case. If, for example, you stand on a very hard, horizontal surface such as a cement pavement, your feet exert on it a downward force equal to your weight; nevertheless, the surface itself suffers no noticeable deformation or acceleration.

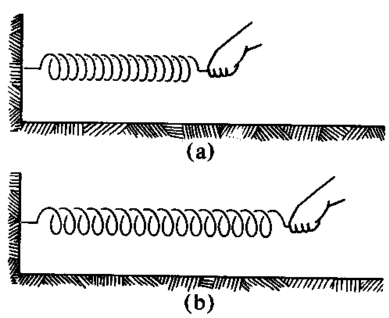


Figure 4-1

One very important property of force is that it has associated with it both a magnitude and a sense of direction. That is, if an agent exerts a force on a body, then he can push or pull on it with various degrees of intensity and in various directions, and the observed deformation or acceleration of the body will in general be different, depending on his choice. This suggests, as will be confirmed below, that force is a vector quantity. However, simply because a physical quantity has associated with it both a magnitude and a direction does not always imply that that physical quantity is a vector. One of the very important and decisive properties of vectors is the rule for the addition of two vectors as defined in Chapter 3. Hence, to establish that force is a vector quantity, it is necessary to define the concept precisely and then to confirm by use of this definition that the effect produced by the simultaneous exertion of two or more forces is the same as that produced by their vector sum.

For many purposes it is convenient to divide the various forces that can act on a material body into two classes. In one of these belong those forces which arise from other bodies in physical contact with the one in question; these are known as *contact forces.* If, for example, you place your hand against a wall and push, then, since you are in contact with the wall, the force you exert on it is a contact force. In the second class belong those forces which arise from matter *not* in contact with the body in question; and these are known as *action-at-a-distance forces.* Examples of the latter are: gravitational forces; electrical forces, which act between bodies carrying an electric charge; and nuclear forces, which act between the nuclei of atoms. Since nuclear forces are of very short range in that they are effective only over distances of the order of 10^{-15} meter, they are of no interest at the moment. On the other hand, since we are residents of the rather massive planet earth, gravitational forces play a very important role in our lives, and we shall discuss the effects of gravity in some detail below. Also of considerable importance to us are electric forces. These are of interest not only in connection with our studies of electromagnetic phenomena but also because *contact forces are actually the manifestation on a macroscopic scale of such electric forces.* That is, since the forces between atoms and molecules are basically electrical in nature and since the forces between bodies in contact are due to molecular interactions, it follows that contact forces are basically of the action-at-a-distance type. Strictly speaking, therefore, most forces that we know of are of the action-at-a-distance type, but for purposes of discussing macroscopic phenomena, it is convenient to continue to think of them as belonging to one or the other of these two classes.

This distinction between action-at-a-distance forces and contact forces is illustrated in Figure 4-2a, which shows a man of weight w standing on a horizontal surface. Figure 4-2b shows the forces acting on him. First, there is the force of gravity, which is of the action-at-a-distance type; it acts vertically downward and has a strength numerically equal to his weight w. The second force acting is a normal force **V** which the floor, with which he is in contact, exerts on him. This normal force **V**, as we shall see later, must be equal and opposite to the downward force **w**. Are there any other forces acting on the man? Since he has no excess electric charge there are none of the action-at-a-distance variety, and since he is not in physical contact with any other bodies there are no additional contact forces acting on him either. It

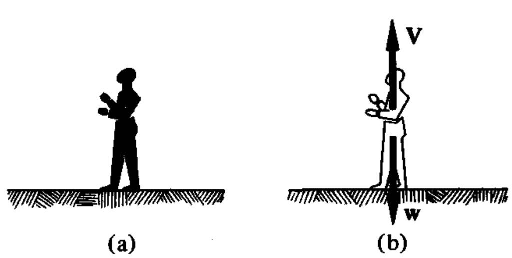


Figure 4-2

follows that the two forces labeled **V** and **w** in the figure are the only ones acting on him.

4-3 Newton's first law

Imagine placing a flat object on a rough, horizontal surface and giving it an impulsive push so that it slides along the surface. Experiment shows that the object continually slows down from its initial velocity, and after traveling a certain distance along the surface it comes to rest. Since its velocity constantly changes during this slowing-down process, it follows that during its motion the object accelerates in a direction opposite to its velocity, and we conclude that a force must be acting on it. We call this particular force that tends to oppose the relative sliding motion of two surfaces in contact a *frictional force.*

Imagine now repeating the above experiment over and over, but with sucessively smoother surfaces, obtained, say, by polishing them. For the same initial push each time, we find that the smoother the surfaces are, the farther the block slides before coming to rest. That is, the smoother the two surfaces in contact are, the less the object accelerates and the weaker must be the force of friction acting on it. Presumably, then, in the limit as the surfaces become "infinitely" smooth, the force of friction vanishes completely and the block will not slow down at all.

Newton's first law is a statement about the idealized limit of such experiments and was originally stated by Newton approximately as follows:

> *Every body continues in its state of rest or of uniform motion in a straight line unless it is compelled by forces to change that state.*

Since being at rest or in uniform motion in a straight line is another way of saying zero acceleration, this law says in effect that if there are no forces acting on a body then that body cannot accelerate. And, conversely, this law can also be thought of as saying that if a body accelerates—that is, it is neither at rest nor moving uniformly in a straight line—then the forces acting on it are *not* zero. Note that even though the concept of force has not been precisely defined, this law is still useful since it expresses in operational terms what is meant by the term "zero force." This first law of Newton is also known as *Galileo's principle of inertia.*

In our studies of kinematics we saw that the velocity and acceleration of a particle will in general be different for two observers in relative motion. Thus, if you see an automobile accelerating north, then to the driver of the automobile, *you* are accelerating south. Since Newton's first law deals with the acceleration of a particle, and since this acceleration will in general be different for observers in relative motion, it follows that this law can be valid

only in selected coordinate systems. An *inertial coordinate system or reference frame* is defined as one in which Newton's first law is valid. Thus an inertial reference frame is one in which a body not acted on by any forces will continue in its state of rest, or of uniform motion in a straight line. Newton's first law, in other words, can be thought of as defining what is meant by an inertial coordinate system. The term *Newtonian reference frame* is sometimes used in place of the term inertial reference frame.

If a particle is moving at a uniform velocity with respect to an observer O, then it will also appear to move with a uniform velocity relative to a second observer O', provided that his velocity relative to O is constant (see Section 2-9). Hence, all coordinate systems that move with uniform velocity relative to an inertial reference frame are themselves inertial reference frames. Sometimes, in defiance of the fact that the stars are actually moving, an inertial coordinate system is said to be one which is at rest relative to the "fixed stars." In this language, any system that moves with uniform velocity relative to the fixed stars is an inertial system.

Are we observers fixed in an inertial reference frame? Because of the earth's rotation about its axis and its motion about the sun, both of which involve centripetal acceleration, strictly speaking, the answer is no. However, the accelerations involved in these terrestrial motions are very small ($\leqslant 3 \times 10^{-2}$ m/s^2) and will usually be neglected. For these situations, then, a person stationary with respect to the earth will be thought of as being in an inertial reference frame. Strictly speaking, only astronauts, who are traveling out in space with uniform velocity relative to the fixed stars are stationary relative to an inertial frame.

Because of the existence of gravitational and frictional forces, care must be exercised in the laboratory in carrying out experiments designed to test the validity of Newton's first law. Figure 4-3 shows a possible setup. If an object is placed on a horizontal air track, it will be supported by a thin layer of air, which is pumped through the small holes on the upper surface of the track. Because of this film of air, the frictional force between the two surfaces is generally very small, and can be safely neglected. If, in addition, the track is horizontal, gravity will not affect the motion of the object on the air track either. Under these conditions, then, the object is free to slide along the track without any forces acting along the direction of motion. According to the first law, it should slide along the track with uniform velocity; this is consistent with results obtained by careful laboratory experiments.

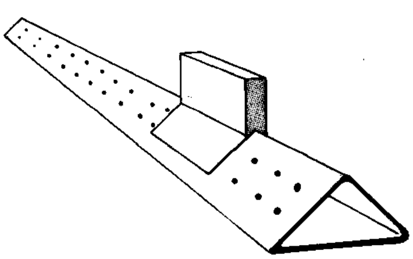


Figure 4-3

4-4 Mass

If two similar bodies are subjected to precisely the same force, then depending on the material of the bodies they may undergo very different accelerations. Thus, two solid objects of the same size, but one made of wood and the other of lead, will not have identical accelerations if pushed by the same force. We attribute this difference in behavior of the two bodies to a property of matter called *inertia* and we define *mass* to be a measure of inertia. The greater the inertia of a body, the less it will accelerate when subjected to a given force.

As an approximation to the idealistic situation of two bodies in an inertial frame, consider in Figure 4-4 two bodies, 1 and 2, which are mounted on a horizontal air track and free to move only along the direction of the track. Under these circumstances, with neither gravity nor friction involved, the only forces that can affect the one-dimensional motion of the bodies are those they exert on each other. Suppose now that we allow these bodies to interact successively through a variety of forces, and observe their associated accelerations. For example, each of the bodies may be given an electric charge so that they can interact via electric forces, or a small compressed spring may be placed between them so that they are pushed apart or, if possible, they could be magnetized so that they influence each other's motion magnetically. If $\mathbf{a}_1$ and $\mathbf{a}_2$ are the respective accelerations of bodies 1 and 2 associated with *any* given choice for such a force, then experiment shows that $\mathbf{a}_1$ and $\mathbf{a}_2$ are always oriented in opposite directions and that the ratio of their magnitudes is a *time-independent constant.* We may summarize the results of these experiments by

$$\mathbf{a}_1 = -k_{12}\mathbf{a}_2 \tag{4-1}$$

with k_{12} a positive, time-independent constant. If a second experiment is carried out with these same two bodies but this time with a different choice of force, then the resultant accelerations $\mathbf{a}_1'$ and $\mathbf{a}_2'$ are again related by

$$\mathbf{a}_1' = -k_{12}\mathbf{a}_2'$$

with k_{12} precisely the same constant as in (4-1). In other words, (4-1) is *independent of the nature of the forces acting between the bodies.* Hence no

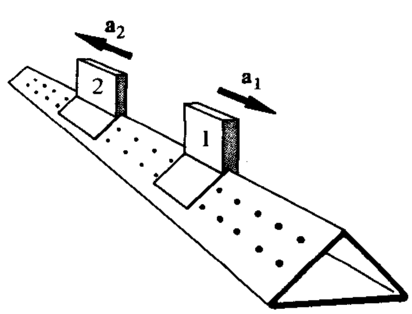


Figure 4-4

matter how the two bodies in Figure 4-4 are allowed to interact, the ratio of the magnitudes of their accelerations has always the same value k_{12}. It follows therefore that the parameter k_{12} as defined in (4-1) is a property only of the two bodies 1 and 2 and is in no way dependent on the nature of the force acting between them.

Let us now introduce a third body, call it 3, and carry out the above experiments first with 1 and 3 as the objects and then with 2 and 3 as the objects. As before, these experiments show that when the bodies are subject to various interactions, the relation between their accelerations may be characterized by certain constants, which we shall call k_{13} and k_{23}, respectively. Specifically, the experiments with 1 and 3 lead to

$$\mathbf{a}_1 = -k_{13}\mathbf{a}_3$$

where $\mathbf{a}_1$ and $\mathbf{a}_3$ are the respective accelerations of 1 and 3 and where k_{13} is a constant characteristic only of these two bodies. In a similar way, by utilizing bodies 2 and 3 we obtain a value of the constant k_{23}. Comparison of these experimental values for k_{12}, k_{13}, and k_{23} shows that they are related by

$$k_{23} = \frac{k_{13}}{k_{12}} \tag{4-2}$$

Moreover, since the three bodies are arbitrary, this relation is valid for any triplet of bodies.

Now as noted above, k_{12} has to do only with the properties of bodies 1 and 2, and similarly k_{13} and k_{23} depend only on the properties of 1 and 3 and of 2 and 3, respectively. It follows then from (4-2) that since the left-hand side depends only on the properties of 2 and 3, that the ratio k_{13}/k_{12} *must be independent of the properties of body* 1. That is, even though k_{13} depends on the properties of 1 and 3, and k_{12} depends on those of 1 and 2, their ratio is independent of the characteristics of body 1. From this feature and the fact that by definition of the constant k_{12}, (4-1) implies that $k_{12} = 1/k_{21}$, it follows that k_{12}, k_{13}, k_{23} may be expressed as the ratios

$$k_{12} = \frac{m_2}{m_1} \qquad k_{13} = \frac{m_3}{m_1} \qquad k_{23} = \frac{m_3}{m_2} \tag{4-3}$$

with m_1 depending only on the properties of 1, and similarly for bodies 2 and 3. It is easy to confirm that (4-3) satisfies the experimental result in (4-2) and the conditions $k_{12} = 1/k_{21}$ and so on. These three quantities m_1, m_2 and m_3 are defined to be masses of bodies 1, 2, and 3, respectively. In terms of these masses, the experimental result in (4-1) may be expressed in the equivalent form

$$m_1\mathbf{a}_1 = -m_2\mathbf{a}_2 \tag{4-4}$$

It is important to note that by use of (4-1) and (4-3) we have available an operational procedure for measuring the ratio of the masses of any two bodies. Thus, to give a meaning to the mass of any body, it suffices to define a

standard in terms of which this mass may be expressed. As discussed in Chapter 1, the standard universally accepted today is the platinum iridium cylinder which is housed at the Bureau of Weights and Measures in Sèvres, France. It defines the kilogram. In principle, making use of this standard as one of the two bodies in an experiment of the above type, the mass of any body may be expressed in terms of that of the standard. In this way then by measuring acceleration we can assign, in operational terms, a value for the mass of any body.

There is a very important property of mass that also follows from experiments of the above type. Suppose that the measured values of the masses of two bodies are m_1 and m_2. Let us now combine these two bodies to form a third, and carry out measurements of its mass. Experiment shows that the mass M of this combined body has the value

$$M = m_1 + m_2 \tag{4-5}$$

and this is known as the *additivity property* of mass. It is because of this experimental feature—that the mass of a body is the sum of the masses of its constituents—that we often describe mass as a measure of the amount of matter in a body. For scientific work, however, the foregoing operational definition of mass is to be preferred.

4-5 Newton's second law

Consider an isolated body of mass m at rest in an inertial reference frame and suppose that an external agent exerts a push or a pull on the body so that it accelerates. Experience tells us that the greater is the mass of the body the less will be its acceleration and that, for a fixed mass, the stronger the push or pull the larger will be its acceleration. This suggests that we define the force $\mathbf{f}$ acting on a body of mass m by

$$\mathbf{f} = m\mathbf{a} \tag{4-6}$$

where $\mathbf{a}$ is the observed acceleration of the body. Note that (subject to self-consistency checks) $\mathbf{f}$ as defined here is a vector that is oriented along the direction of $\mathbf{a}$.

Since the kilogram is the SI unit of mass and the meter per second squared the unit of acceleration, it follows that the kilogram meter per second squared (kg-m/s^2) is the SI unit of force. It is convenient to introduce the unit of force of the newton (N) by the relation

$$1\ \mathrm{N} = 1\ \frac{\text{kg-m}}{\mathrm{s}^2}$$

so that a newton is the force that when applied to a 1-kg body will give it an acceleration of 1 m/s^2.

Since the acceleration $\mathbf{a}$ of a particle is a vector, it follows from (4-6) that $\mathbf{f}$ must be one also. In order for this definition of force to be consistent it is

necessary that the sum of two forces be the vector obtained by use of the rules of vector addition. To confirm the fact that this is indeed the case, suppose that in Figure 4-5a a body of mass m is subject to a certain force $\mathbf{f}_1$. Then, by use of (4-6), its acceleration $\mathbf{a}_1$ is

$$\mathbf{a}_1 = \frac{1}{m}\mathbf{f}_1$$

Similarly, if it is subjected to a different force $\mathbf{f}_2$, then, as shown in Figure 4-5b its acceleration $\mathbf{a}_2$ is $\mathbf{f}_2/m$. If now the body is subjected to the simultaneous action of both $\mathbf{f}_1$ and $\mathbf{f}_2$ (see Figure 4-5c), then experiment shows that its acceleration $\mathbf{a}$ under these circumstances is the *vector* sum

$$\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$$

It follows that the definition in (4-6) will be self-consistent, provided that the total force $\mathbf{f}$ is interpreted to be the vector sum

$$\mathbf{f} = \mathbf{f}_1 + \mathbf{f}_2$$

This then establishes the fact that force as defined in (4-6) is indeed a vector.

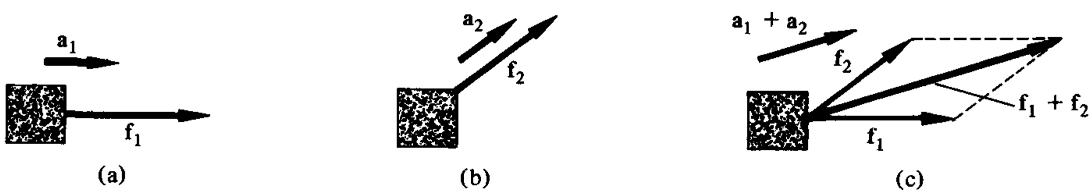


Figure 4-5

Generalizing this experimental result to the case of an arbitrary number of forces we may state *Newton's second law*:

If a body of mass m is subject simultaneously to various forces $\mathbf{f}_1$, $\mathbf{f}_2 \ldots$, *and if* $\mathbf{a}$ *is the acceleration of the body as seen in an inertial reference frame, then*

$$\boxed{\mathbf{F} = m\mathbf{a}} \tag{4-7}$$

where $\mathbf{F}$ *is the net force or the vector sum of the individual forces:*

$$\mathbf{F} = \mathbf{f}_1 + \mathbf{f}_2 + \mathbf{f}_3 + \cdots$$

Note that even though (4-7) appears to be formally identical to (4-6), they are actually different. The latter *defines* a force $\mathbf{f}$ in operational terms, whereas the former relates the vector sum $\mathbf{F}$ of *all* forces acting on the body to its acceleration.

In terms of components, for motion in the x-y plane Newton's second law may be expressed in the form

$$F_x = ma_x \qquad F_y = ma_y \tag{4-8}$$

where F_x is the component of the net or unbalanced force **F** along the x-axis, a_x is the component of the resultant acceleration along the same axis, and similarly for F_y and a_y along the y-axis. In solving problems the component forms in (4-8) are generally the most useful. Note that these relations in (4-8) are valid regardless of how the orientation of the coordinate axes or the origin is selected. For cases in which the force acts in a single direction we shall also use the one-dimensional form of (4-7). This is given by

$$F = ma \tag{4-9}$$

where the symbols F and a refer to the components of the force and acceleration along the direction of action of the force.

For the special case in which the vector sum **F** of all the forces acting on the body vanishes, it follows from the second law that its acceleration must vanish. But zero acceleration means that the body is either at rest or else moving with uniform velocity in a straight line. Hence follows the compatibility of Newton's second law with the first law.

Example 4-1 A block having a mass of 20 grams is pulled along a smooth, horizontal surface with a constant, horizontal force of 0.5 newton. What is the acceleration of the block?

Solution Since the surface is smooth, it follows that there are no frictional forces acting, and since it is horizontal, the force of gravity does not play a role either. Thus only the given force of 0.5 newton acts, and it follows that the acceleration is along the direction of this force. Substituting the given values for F and m into (4-9), we find that

$$a = \frac{F}{m} = \frac{0.5\ \text{N}}{0.02\ \text{kg}} = 25\ \text{m/s}^2$$

since there are 1000 grams in 1 kg. Note that in making use of (4-9) it is necessary that F be expressed in newtons, m in kilograms, and a in meters per second squared.

Example 4-2 A body of mass m moves along the x-axis of a certain coordinate system so that at time t its position is

$$x(t) = \alpha t^3 - \beta t$$

with α and β positive constants.

(a) Calculate the acceleration of the body.

(b) What is the force acting on it?

Solution

(a) Making use of the rules for differentiation, we find that its velocity v_x in the x-direction is

$$v_x = \frac{dx}{dt} = \frac{d}{dt}(\alpha t^3 - \beta t) = 3\alpha t^2 - \beta$$

and a second differentiation then gives the acceleration a_x, as follows:

$$a_x = \frac{dv_x}{dt} = \frac{d}{dt}(3\alpha t^2 - \beta) = 6\alpha t$$

(b) Substituting this form for a_x into (4-8), we obtain

$$F_x = ma_x = 6\alpha mt$$

where m is the mass of the body. Since there is no motion along the y-direction, there is no acceleration along this axis either. Hence the component F_y of $\mathbf{F}$ along the y-direction is zero.

Example 4-3 Figure 4-6 shows a block of mass 2 kg acted upon by the two forces $\mathbf{F}_1 = \mathbf{i}2$ newtons and $\mathbf{F}_2 = \mathbf{j}1$ newton, with $\mathbf{i}$ and $\mathbf{j}$ unit vectors along the appropriate axes. Calculate the acceleration of the block.

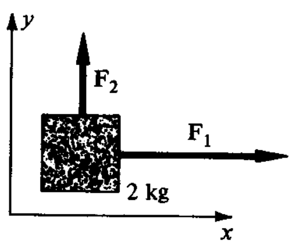


Figure 4-6

Solution Making use of the unit vectors $\mathbf{i}$ and $\mathbf{j}$, we may express the total force $\mathbf{F}$ acting on the block in the form

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 = (2\mathbf{i} + \mathbf{j})\ \mathrm{N}$$

or, equivalently,

$$F_x = 2\ \mathrm{N} \qquad F_y = 1\ \mathrm{N}$$

Substituting these values into (4-8), we obtain

$$a_x = \frac{F_x}{m} = \frac{2\ \mathrm{N}}{2\ \mathrm{kg}} = 1\ \mathrm{m/s^2}$$

$$a_y = \frac{F_y}{m} = \frac{1\ \mathrm{N}}{2\ \mathrm{kg}} = 0.5\ \mathrm{m/s^2}$$

and this leads to

$$\mathbf{a} = (\mathbf{i} + 0.5\mathbf{j})\ \mathrm{m/s^2}$$

Example 4-4 A truck of mass 6000 kg accelerates from rest with constant acceleration a_0 to 30 km/hr in 5.0 seconds. What force is required to do this?

Solution Since the acceleration a_0 is constant, (2-11) is applicable. Solving this formula for a_0 and substituting the given values $v_0 = 0$, $t = 5.0$ seconds, $v(5\ \mathrm{s}) =$ 30 km/hr we find

$$a_0 = \frac{v(5\ \mathrm{s})}{5.0\ \mathrm{s}} = \frac{1}{5.0\ \mathrm{s}} \times 30\,\frac{\mathrm{km}}{\mathrm{hr}} \times \frac{1\ \mathrm{hr}}{3600\ \mathrm{s}} \times \frac{10^3\ \mathrm{m}}{\mathrm{km}} = 1.7\ \mathrm{m/s^2}$$

Recognizing that the force acts only along a single direction we obtain by the use of (4-9),

$$\begin{aligned} F &= ma_0 = 6.0 \times 10^3\ \mathrm{kg} \times 1.7\ \mathrm{m/s^2} \\ &= 1.0 \times 10^4\ \mathrm{N} \end{aligned}$$

This horizontal force is produced on the truck by the road, which pushes on the tires and thus forces the truck to accelerate.

Example 4-5 Assuming that the earth ($m = 6.0 \times 10^{24}$ kg) orbits with uniform speed about the sun in a circular path of radius $R = 1.5 \times 10^8$ km, calculate the force required to sustain this motion (see Figure 4-7).

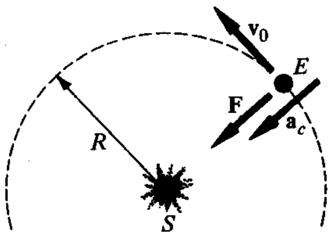


Figure 4-7

Solution As we saw in Chapter 3, uniform motion in a circle is invariably accompanied by centripetal acceleration, which is directed from the moving body to the center of the circle.

Since there are approximately 3.16×10^7 seconds in a year it follows that the orbital speed v_0 of the earth is

$$v_0 = \frac{2\pi R}{1\text{ yr}} = \frac{2\pi \times 1.5 \times 10^8\text{ km}}{3.16 \times 10^7\text{ s}}$$
$$= 3.0 \times 10^4\text{ m/s}$$

The magnitude a_c of the centripetal acceleration is thus

$$a_c = \frac{v_0^2}{R} = \frac{(3.0 \times 10^4\text{ m/s})^2}{1.5 \times 10^{11}\text{ m}} = 6.0 \times 10^{-3}\text{ m/s}^2$$

Substituting this value into (4-9), we find for F

$$F = ma_c = 6.0 \times 10^{24}\text{ kg} \times 6.0 \times 10^{-3}\text{ m/s}^2$$
$$= 3.6 \times 10^{22}\text{ N}$$

The direction of this force is shown in the figure to be radially inward from the earth toward the sun. It corresponds to the force of gravitational attraction with which the sun pulls on the earth and causes it to orbit around the sun. It is interesting to note that even though this acceleration of 6.0×10^{-3} m/s^2 is rather small, because of the enormous mass of the sun, the force itself is quite large.

4-6 Newton's third law

In addition to the first and second laws of motion, in the *Principia Mathematica* Newton also enuniciated another law, which we call *Newton's third law.* Unlike the first two laws, this third law is concerned not with the motion of a particle but rather with the relation between the forces that two bodies exert on each other. This law plays a very important role in the analysis of mechanical systems.

In words, the third law is conventionally stated in the form:

To every action there is an equal and opposite reaction.

This means that if body A exerts a force on a second body B, then B reacts by exerting on A a force of equal magnitude but of the opposite orientation. If, for example, a man leans against a wall, thereby exerting on it a certain force directed away from himself, then according to the third law the wall pushes back (reacts) on the man with an equal and opposite force.

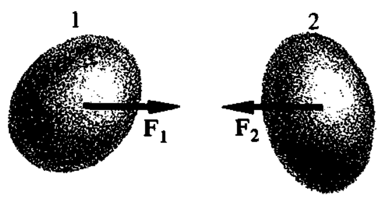


Figure 4-8

Consider, in Figure 4-8, two bodies 1 and 2 and suppose that $\mathbf{F}_1$ is the force on 1 due to 2 and $\mathbf{F}_2$ the force on body 2 due to 1. These forces $\mathbf{F}_1$ and $\mathbf{F}_2$ may be either contact forces or action-at-a-distance forces, or some combination of these. Regardless of what additional forces may be exerted on 1 and 2 by any other nearby bodies, the third law tells us that $\mathbf{F}_1$ and $\mathbf{F}_2$ must be equal and opposite. That is,

$$\boxed{\mathbf{F}_1 = -\mathbf{F}_2} \tag{4-10}$$

As a special case, consider again the situation in Figure 4-4 of two bodies which are effectively isolated from all other matter and for which the only force on each body is due to the proximity of the other. Applying Newton's second law, we have

$$\mathbf{F}_1 = m_1\mathbf{a}_1 \qquad \mathbf{F}_2 = m_2\mathbf{a}_2 \tag{4-11}$$

where $\mathbf{F}_1$ is the force on 1 due to 2 and $\mathbf{a}_1$ is its acceleration, and similarly for $\mathbf{F}_2$ and $\mathbf{a}_2$. Note that (4-11) would *not* be correct if any additional forces acted on the bodies. Comparison of (4-11) with (4-4) leads directly to

$$\mathbf{F}_1 = -\mathbf{F}_2$$

In other words, these forces are equal and opposite. It should be emphasized that the statement of the third law in (4-10) is more general than this, for it allows for the possibility of forces other than $\mathbf{F}_1$ and $\mathbf{F}_2$ to act also.

As a simple application of the third law, consider in Figure 4-9a a block of weight $\mathbf{w}$ resting on a horizontal floor. In Figure 4-9b we have isolated the block and drawn a downward arrow to represent the force of gravity on it (that is, its weight $\mathbf{w}$) and an upward arrow labeled $\mathbf{V}$ to represent the upward push on it by the floor. Since the block is not moving, its acceleration vanishes and so must the total force acting on it. Hence

$$\mathbf{V} = -\mathbf{w} \tag{4-12}$$

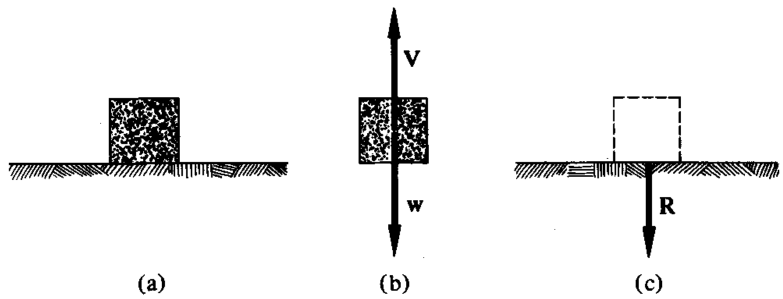


Figure 4-9

or, in other words, the upward push on the block is precisely the same as is the downward pull of gravity.

Consider now, in Figure 4-9c, the floor on which the block rests. Since the block is in contact with the floor, it exerts on it a certain downward force, which has been indicated by a downward arrow labeled **R**. (The various other forces that must act on the floor to keep it from accelerating have not been drawn.) Now, considering the floor and the block as two mutually interacting bodies, the third law tells us that the upward force **V** on the block by the floor must be the same as the downward reaction force on the floor by the block. It follows therefore that $\mathbf{R} = -\mathbf{V}$ or, equivalently, by use of (4-12), that

$$\mathbf{R} = \mathbf{w}$$

Similarly, if the earth and the block are taken to be two mutually interacting bodies, then since the earth pulls down on the block with a force equal to the block's weight it follows by use of the third law that the *block pulls up on the earth with a force of the same strength w.*

A second application of the third law is shown in Figure 4-10a, and involves a block of mass M pulled along a smooth, horizontal surface by a force **F**, which pulls on a massless rope attached to the block. Figure 4-10b shows the isolated rope and the forces acting on it. In addition to the external force **F** acting to the right, the rope also experiences a force **P**, which acts to the left and is the contact force due to the block. By hypothesis the rope has zero mass, and thus it follows from the *second law* (not the third) that the net force acting on the rope vanishes:

$$\mathbf{P} = -\mathbf{F} \tag{4-13}$$

Continuing with the analysis, the arrow in Figure 4-10c represents the force **Q** with which the rope pulls on the block. If we consider the block and the rope to comprise a mutually interacting pair, then by use of the third law it follows that the force on the block by the rope must be equal and opposite to the force

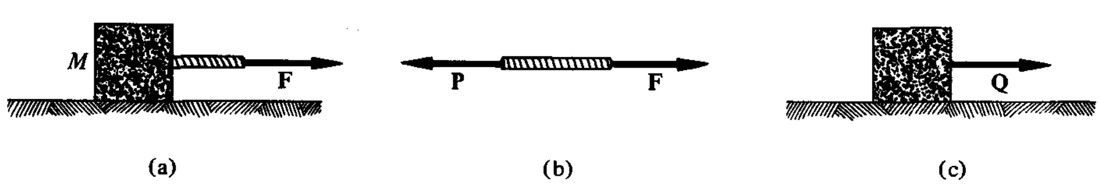


Figure 4-10

with which the block pulls the rope. Hence it follows from the third law that $\mathbf{Q} = -\mathbf{P}$, and combining this with (4-13) and using the second law, we conclude that

$$\mathbf{F} = M\mathbf{a} \tag{4-14}$$

with $\mathbf{a}$ the observed acceleration of the block. In effect, then, because of the fact that the rope is massless, this result is the same as if we had applied the force to the block directly. In the problems the analogues of (4-13) and (4-14) for the case in which the mass m of the rope is not negligible are established.

As a final application of the third law let us consider an experiment described by Newton in partial justification for this law. It had been generally known at that time that an ordinary magnet would attract iron filings, but it was not equally appreciated that the iron filings would *also attract a magnet.* Newton carried out an experiment by placing a magnet in one vessel, some iron filings in another, and, as shown in Figure 4-11, floating both vessels in water. By restricting the motion of the vessel containing the filings he confirmed that the iron filings also attract the magnet. Furthermore, he found that if he placed the two floating vessels in physical contact, then neither one could propel the other through the water. That is to say, he showed that consistent with the third law the force with which the iron filings attract the magnet is precisely equal and opposite to the force with which the magnet attracts the filings. For otherwise, and in contradiction with experiment, the composite system would have to accelerate in accordance with the second law.

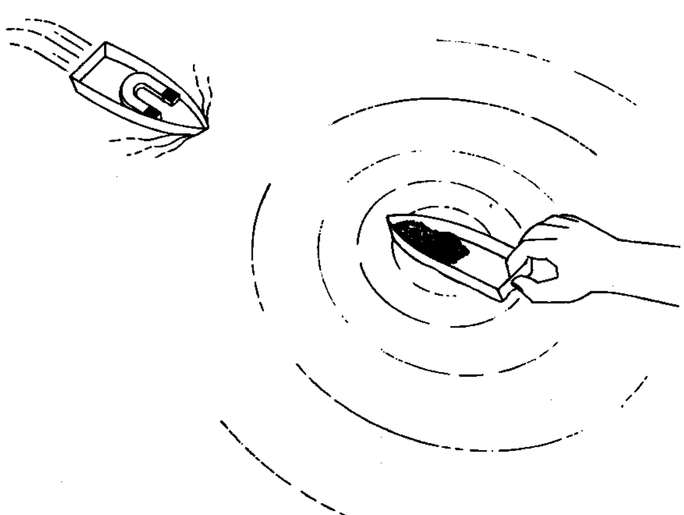


Figure 4-11

4-7 Newton's law of universal gravitation

One of the very important forces between any two bodies, particularly if one of them is as massive as is our earth, is the force of gravity. In this section we shall describe this force.

In the famous *Principia Mathematica* in which Sir Isaac Newton presented his three laws of motion, he also described his very profound law of *universal gravitation.* This law states:

Every body in the universe attracts every other body with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

Figure 4-12 shows two bodies of masses m_1 and m_2 and separated by a distance r. According to the law of universal gravitation, the force **F** with which body 2 attracts body 1 is directed from 1 to 2, and has a magnitude

$$F = \frac{Gm_1m_2}{r^2} \tag{4-15}$$

where G is a coefficient of proportionality, known as the *gravitational constant.* Similarly, the *magnitude* of the force that 1 exerts on 2 is also given by (4-15), but as shown in the figure it is oriented in the opposite direction.

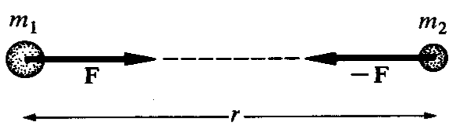


Figure 4-12

The numerical value of the gravitational constant G is of considerable interest, and plays an important role in all considerations of gravitational effects. The first person to succeed in its measurement was Henry Cavendish (1731–1810), who obtained the value 6.75×10^{-11} N-m^2/kg^2. This is surprisingly accurate and is within a few percent of the presently accepted value of

$$G = 6.6732 \times 10^{-11} \frac{\text{N-m}^2}{\text{kg}^2} \tag{4-16}$$

Figure 4-13 illustrates the principle underlying the Cavendish experiment. Two small spheres a and a', each of mass m, are suspended from the ends bb' of a very light and horizontal rod, which is itself suspended at its midpoint by a very strong fiber cc'. If we now bring up to a and a' two very massive spheres A and A', each of mass M, then because of the gravitational attraction between A' and a' and between A and a, the rod will rotate slightly about cc', thus causing the fiber to twist. The angle through which the fiber is twisted is then a measure of the gravitational force F between the two pairs of spheres. In terms of (4-15), this means that F is known. Since m, M, and r can each be measured independently, it follows that in effect we are able to measure G. Because of the very small value for G, the angle of twist of the fiber is also generally very small, and thus extreme care must be used in its measurement. As shown in the figure, this small angle can be measured by attaching a

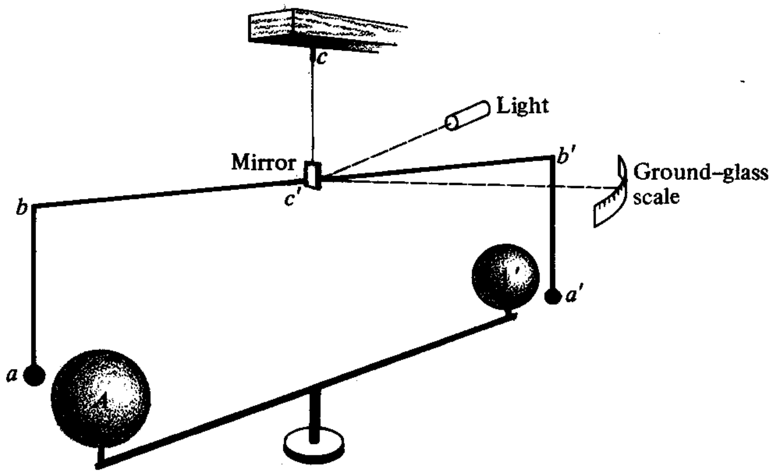


Figure 4-13

mirror to the rod bb' and allowing light from a source to fall on this mirror. Any rotations about cc' can then be measured by observing the displacement of the reflected beam on a ground-glass scale.

Consider an object of mass m at a point on the surface of the earth. Assuming the earth to be a sphere of radius R it follows,[1] by use of (4-15), that the force F on the body is

$$F = \frac{GmM_E}{R^2} \tag{4-17}$$

where $M_E = 6.0 \times 10^{24}$ kg is the mass of the earth. The *weight*, w, of the body is, by definition, this force F. Thus we may write by use of (4-17)

$$w = mg \tag{4-18}$$

with g, the *acceleration of gravity*, defined by

$$g = \frac{GM_E}{R^2} \tag{4-19}$$

The weight w in (4-18) represents only the magnitude of the gravitational force on the body; the direction of this force is vertically downward, from the body to the center of the earth.

Making use of the known value of 6.0×10^{24} kg for M_E and the value 6.4×10^3 km for the radius R of the earth, we can easily obtain a numerical value for g by use of (4-19). At sea level, at the equator we find

$$g = 9.8 \text{ m/s}^2 \tag{4-20}$$

By the very nature of its definition, the numerical value of g will vary slightly from place to place on earth because of variations in elevation, the local value

[1]As will be evident from the analogous discussion of the electric field in Chapter 20, the gravitational force exerted by a spherically symmetric body is the same as if all of the mass of the body were concentrated at its center.

of the earth's density, and so forth. In addition, we shall see later that g also varies with latitude because of the earth's rotation. Generally speaking, however, these effects are quite small, and amount to less than 1 percent. As such they can be safely neglected.

Although the weight of an object has been defined when it is on earth, (4-19) can also be used for other planets by defining an appropriate value for g. Thus the weight w_M of an object of mass m when it is on the surface of the moon is

$$w_M = mg_M$$

where g_M is the gravitational acceleration on the moon and is defined by

$$g = \frac{GM_M}{R_M^2} \tag{4-21}$$

with M_M the mass and R_M the radius of the moon. Using the known values $M_M = 7.3 \times 10^{22}$ kg, and $R_M = 1.74 \times 10^6$ meters, we find for g_M the value

$$g_M = 1.6 \text{ m/s}^2$$

and this corresponds to a gravitational acceleration of about one sixth of the terrestrial value in (4-20). It follows, therefore, that if w is the weight of an object on earth, the corresponding weight when it is on the lunar surface is $w/6$.

In Appendix B will be found a table of planetary data that can be used to calculate the g-value on other planets.

Example 4-6 What is the gravitational attraction between two protons in a nucleus of helium, where they are separated by a distance of about 10^{-15} meter?

Solution Using the known value $m = 1.67 \times 10^{-27}$ kg we find on substituting into (4-15) that

$$F = G\frac{m_1 m_2}{r^2} = 6.67 \times 10^{-11} \frac{\text{N-m}^2}{\text{kg}^2} \times \frac{(1.67 \times 10^{-27} \text{ kg})^2}{(10^{-15} \text{ m})^2}$$
$$= 1.9 \times 10^{-34} \text{ N}$$

which is a very small force indeed.

Example 4-7 Calculate the grávitational attraction between two people, each of mass 100 kg and compare this with their weight. Assume they are separated by a distance of 2 meters.

Solution Substituting the given data into (4-15), we obtain, for the force between them,

$$F = \frac{Gm_1 m_2}{r^2} = 6.67 \times 10^{-11} \frac{\text{N-m}^2}{\text{kg}^2} \times \frac{(100 \text{ kg})^2}{(2\text{m})^2}$$
$$= 1.7 \times 10^{-7} \text{ N}$$

According to (4-18), the weight w of a 100-kg person is

$$w = mg = 100 \text{ kg} \times 9.8 \text{ m/s}^2 = 980 \text{ N}$$

Thus the weight of either person is about 10^{10} times greater than is the gravitational attraction between them!

Example 4-8 The mass of the planet Jupiter is about 318 times that of the earth and its diameter is 11 times greater. What is the weight on this planet of an astronaut whose weight on earth is 600 newtons?

Solution According to (4-19), the ratio of the g-value g_J on Jupiter to that on earth is

$$\frac{g_J}{g} = \left(\frac{GM_J}{R_J^2}\right)\left(\frac{R_E^2}{GM_E}\right) = \left(\frac{M_J}{M_E}\right) \times \left(\frac{R_E}{R_J}\right)^2$$
$$= \frac{318}{11^2} = 2.6$$

Therefore the astronaut's weight w_J on Jupiter is 2.6 times that on earth, or 2.6×600 newtons $= 1.6 \times 10^3$ newtons.

4-8 Further applications

This section presents some additional examples illustrating Newton's laws.

Example 4-9 Suppose the blocks A and B shown in Figure 4-14a have the respective mass $m_1 = 1$ kg and $m_2 = 3$ kg, and are on a smooth, horizontal surface. If a horizontal force F of 3 newtons pushes them, calculate the acceleration of the system and the force that the 1 kg block exerts on the other block.

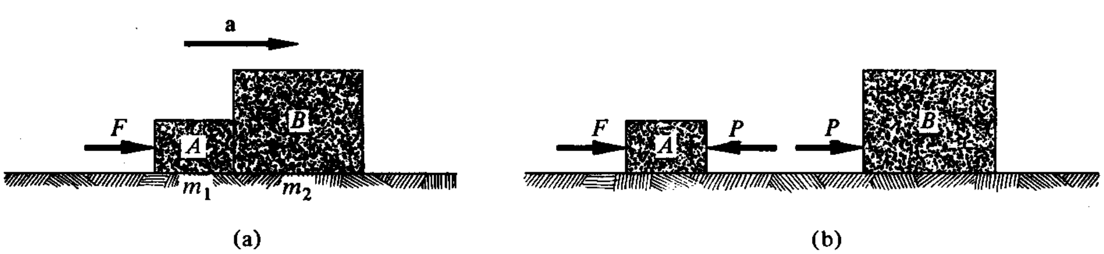


Figure 4-14

Solution Figure 4-14b shows the two blocks isolated and the various forces acting on them. On A there is the force F acting to the right and the unknown force P that B exerts on A, and which has been assumed to be acting to the left. On B there is only the single force P that A exerts on B, and which in accordance with the third law must point to the right. Assuming that the positive axis of a coordinate system points to the right, the equations of motion for A and B are

$$F - P = m_1 a$$
$$P = m_2 a$$

since both bodies undergo the same acceleration a. If we add these two equations, the unknown P drops out. Solving the result for a, we find that

$$a = \frac{F}{m_1 + m_2} = \frac{3\ \text{N}}{1\ \text{kg} + 3\ \text{kg}} = 0.75\ \text{m/s}^2$$

Substituting this back into either of the foregoing relations, we find for the force P that the blocks exert on each other the value

$$P = 2.25\ \text{N}$$

It is important to note the role played by the third law in this derivation. Also of interest is the fact that if A and B are thought of as a single body of mass 4 kg, we obtain the same value as above for the acceleration, namely $a = F/m = 3\ \text{N}/4\ \text{kg} = 0.75\ \text{m/s}^2$. However, to calculate the "internal force" P that the blocks exert on each other, it is necessary to use the more complex procedure.

Example 4-10 Figure 4-15a shows a body of mass m_1 lying on a smooth, horizontal surface and connected by a massless rope, which hangs over a smooth, massless pulley to a second body of mass m_2. Find the acceleration of the bodies and the stretching force or *tension* T in the string.

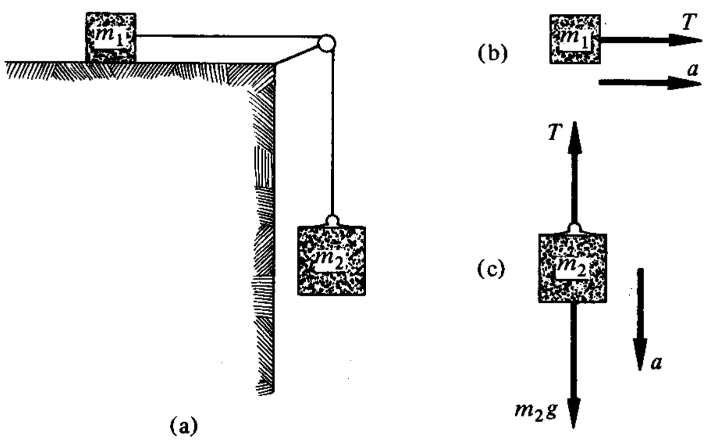


Figure 4-15

Solution Let us apply the second law separately to each body. Since the table is smooth and horizontal it follows, as shown in Figure 4-15b, that only the tension T in the string is effective in producing a horizontal acceleration a to the right. Hence, by the second law,

$$T = m_1 a$$

Figure 4-15c shows the two forces acting on the other body. These are the force $m_2 g$ directed vertically downward and the tension T directed upward. Since the direction of the acceleration has been taken to be positive downward, we obtain

$$m_2 g - T = m_2 a$$

Adding these two equations, the unknown tension T cancels and we find that

$$a = \frac{m_2 g}{m_1 + m_2}$$

The substitution of this into either of the above equations yields

$$T = \frac{m_1 m_2}{m_1 + m_2} g$$

Thus if $m_1 = 2$ kg and $m_2 = 3$ kg, then $a = 0.6\,g = 5.9\ \text{m/s}^2$ and $T = 12$ newtons.

Example 4-11 Figure 4-16a shows the essential components of a device known as an *Atwood's machine.* It consists of two bodies of masses m_1 and m_2 and connected by, let us assume, a weightless rope which hangs over a smooth massless and frictionless pulley. Assuming that $m_2 > m_1$, so that m_2 falls while m_1 rises, find the acceleration of the system and the tension T in the rope.

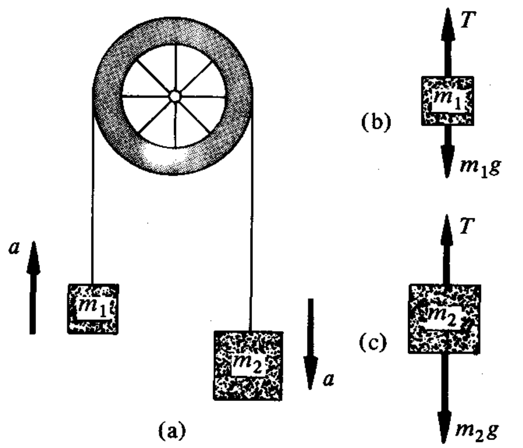


Figure 4-16

Solution Since the pulley is massless and frictionless, its function, just as in the preceding example, is to turn the direction in which the tension in the string acts by 180°. In Figure 4-16b and c we have isolated the bodies and drawn in all of the forces acting on them. Assuming that the acceleration a of m_1 is upward, it follows from the second law that

$$T - m_1 g = m_1 a$$

since $(T - m_1 g)$ is the total force acting on m_1. Similarly, since a is the downward acceleration of m_2,

$$m_2 g - T = m_2 a$$

Adding these two equations, T drops out and we obtain

$$a = \frac{m_2 - m_1}{m_2 + m_1} g$$

The substitution of this value into either of the equations of motion yields

$$T = \frac{2m_1 m_2}{m_1 + m_2} g$$

For the special choice $m_1 = m_2$ this reduces to the expected result $a = 0$ and $T = m_1 g$, as it should.

†4-9 Accelerated reference frames and Galilean invariance

In the foregoing discussion of the second law it was implicitly assumed that the acceleration **a** of the body of interest was defined relative to an inertial

reference frame. The purpose of this section is to study this assumption in more detail and to see what contradictions might arise in its absence.

To introduce the basic ideas, let us first analyze a dynamical system in terms of a coordinate system in which the second law is not valid. To this end, consider in Figure 4-17a a man of mass m standing in an elevator initially at rest. There are two forces acting on the man: his weight $m\mathbf{g}$, directed vertically downward, and an upward contact force $\mathbf{V}$ exerted on him by the floor of the elevator. Since he is not accelerating, it follows from the second law that $\mathbf{V} = -m\mathbf{g}$. Normally, we are not conscious of this upward force $\mathbf{V}$ since it is always with us! Consider now the same situation but suppose this time the elevator has an upward acceleration $\mathbf{a}$. It is a common experience that, as indicated in Figure 4-17b, the magnitude V of the upward force on the man will now be greater than the earth's attractive force, thus giving him a "heavier" feeling. Indeed, if he were standing on a scale, then while the elevator is accelerating upward he would find that its reading, which is a measure of V, would exceed his body weight. Similarly, when the elevator has a downward acceleration, as in Figure 4-17c, he would find the reading on the scale to be *less* than his body weight, thus implying that the upward force $\mathbf{V}$ has a magnitude less than mg this time.

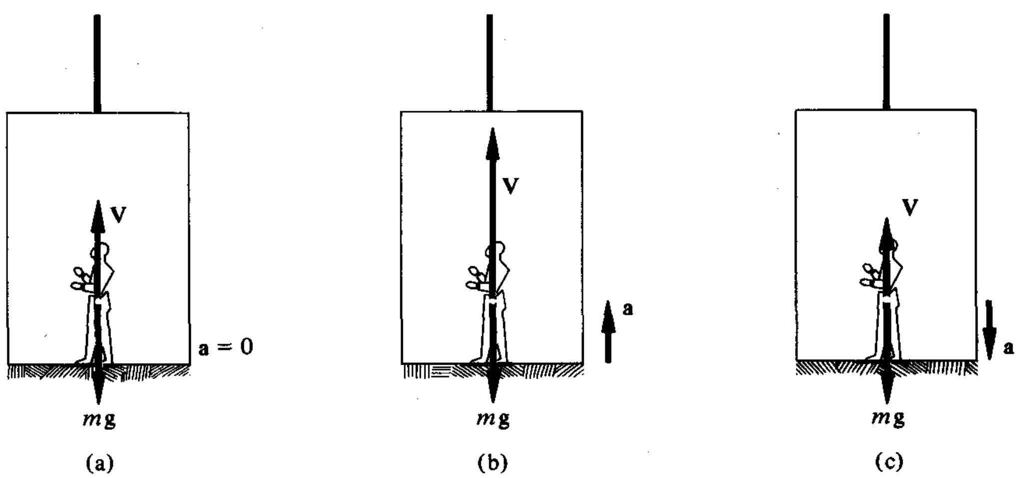


Figure 4-17

At first sight, these two results appear to be contradictory. For if the second law is assumed to be valid in a coordinate system *fixed* with respect to the elevator, then since the man is not accelerating it would follow that the total force acting on him $(\mathbf{V} + m\mathbf{g})$ must vanish. Hence the upward force $\mathbf{V}$ must be equal and opposite to his weight $m\mathbf{g}$. The fact that experimentally this is *not* so means that when the elevator is accelerating up or down, Newton's second law is *not valid* in a system fixed with respect to the elevator.

Note that there is no contradiction whatsoever if we select a coordinate system fixed to the earth. Then, if the elevator has an upward acceleration $\mathbf{a}$, the equation of motion as expressed in the terrestrial frame is

$$\mathbf{V} - m\mathbf{g} = m\mathbf{a}$$

and, consistent with experiment, we find that $V = m(g + a)$. As expected, this is greater than his weight mg. Similarly, when the elevator has a downward acceleration we find that consistent with our experience $V < mg$.

To analyze this possibility of using noninertial reference frames, consider in Figure 4-18 a particle of mass m, which is accelerating under the action of an unbalanced force $\mathbf{F}$ and is viewed by observers in two coordinate systems S and S'. Suppose that S is a *Newtonian frame*—that is, an inertial system in which the second law is valid—and that S' has a uniform *acceleration* $\mathbf{a}_0$ relative to S. We shall only consider the special case for which corresponding axes of the two systems are mutually parallel at all times, so that S' has no rotational motion. Let us assume further that initially the origins of S and S' coincide and that they have a relative velocity $\mathbf{v}_0$ at this instant. It follows that at time t the velocity of S' relative to S is $(\mathbf{v}_0 + \mathbf{a}_0 t)$, and integrating we find that as shown in the figure the displacement of the origin of S' from that of S is $(\mathbf{v}_0 t + \frac{1}{2}\mathbf{a}_0 t^2)$.

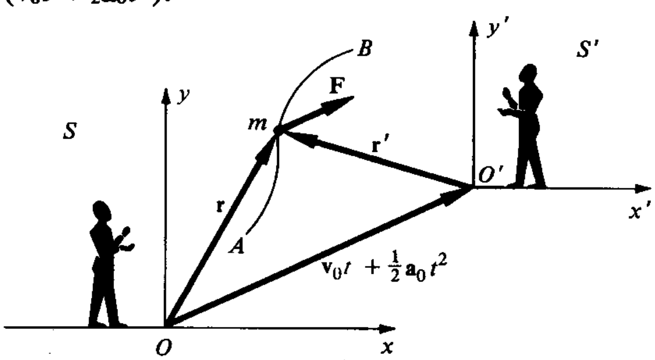


Figure 4-18

Let us now focus attention on the particle which is acted upon by $\mathbf{F}$ and thus travels along the path AB in Figure 4-18. The observer in S describes its position at time t by the vector $\mathbf{r}$ from his origin to the particle, and the observer in S' describes the particle at the same position in space at the same time t by the vector $\mathbf{r}'$ from his origin. The three vectors $\mathbf{r}$, $\mathbf{r}'$, and $(\mathbf{v}_0 t + \frac{1}{2}\mathbf{a}_0 t^2)$ form a closed triangle, and thus

$$\mathbf{r} = \mathbf{r}' + \frac{1}{2}\mathbf{a}_0 t^2 + \mathbf{v}_0 t \tag{4-22}$$

On differentiating (4-22) we obtain

$$\mathbf{v} = \mathbf{v}' + \mathbf{a}_0 t + \mathbf{v}_0 \tag{4-23}$$

where $\mathbf{v} = d\mathbf{r}/dt$ is the velocity of the particle at time t as seen in S and similarly for $\mathbf{v}'$ in S'. Note the important fact that if $\mathbf{a}_0 = 0$, then (4-23) implies that S' will be inertial if S is. In particular, if the velocity $\mathbf{v}$ of the particle in S is constant, then since $\mathbf{v}_0$ is constant and $\mathbf{a}_0 = 0$ it follows that the observed velocity in S' is also constant.

Now, by hypothesis, S is an inertial frame, and thus

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} \tag{4-24}$$

where $\mathbf{F}$ is the total force acting on the body. To obtain the equation of motion as seen by the observer in the accelerating system S', let us substitute for $\mathbf{v}$ by the use of (4-23). The result is

$$\mathbf{F} = m\frac{d}{dt}(\mathbf{v}' + \mathbf{v}_0 + \mathbf{a}_0 t) = m\frac{d\mathbf{v}'}{dt} + m\,\mathbf{a}_0 \tag{4-25}$$

where the final equality follows since $\mathbf{v}_0$ and $\mathbf{a}_0$ are constant. Now the force $\mathbf{F}$ acting on the particle may be thought of in terms of a pointer reading on a scale, and as such it is the same for both observers. Hence, since $d\mathbf{v}'/dt$ is the acceleration of the body as seen in S', it follows that Newton's second law is *not* valid in S'. However, if the term $m\,\mathbf{a}_0$ is transferred to the other side, (4-25) becomes

$$m\frac{d\mathbf{v}'}{dt} = \mathbf{F} - m\,\mathbf{a}_0 \tag{4-26}$$

and this leads us to the main conclusion:

The product of the mass m of a particle and its acceleration as observed in a system S', which has the acceleration $\mathbf{a}_0$ relative to an inertial frame S, is equal to the vector sum of the force $\mathbf{F}$ acting on the particle and the vector $-m\,\mathbf{a}_0$.

For the special case in which $\mathbf{F} = 0$, (4-26) implies that, consistent with our expectations, the observed acceleration of the body is equal and opposite to that of the coordinate system. Thus to the driver of an automobile that is accelerating north at $10\ \mathrm{m/s^2}$ a tree by the roadside appears to accelerate at $10\ \mathrm{m/s^2}$ in a southerly direction.

The fact that the initial velocity $\mathbf{v}_0$ of S' relative to S does not enter (4-26) is of considerable significance. For it implies that if $\mathbf{a}_0 = 0$, then Newton's second law will be valid in both S and S'. This feature that Newton's laws are the same in two systems in relative motion at uniform velocity is usually described by saying that Newton's laws are invariant under *Galilean transformations.*

Now even though the second law is valid only in inertial reference frames, situations frequently do arise for which it is desirable, nevertheless, to make use of an accelerated coordinate system—for example, if the observer is at rest in an accelerating vehicle such as an elevator, an airplane, or a space vehicle. For these situations it is possible to use the second law in noninertial systems, provided that the term $-m\,\mathbf{a}_0$ in (4-26) is included as an additional force. That is, if we assume that the total force acting on the particle is not $\mathbf{F}$ but rather the quantity $(\mathbf{F} - m\,\mathbf{a}_0)$, this force is indeed equal to the product of $(m\,d\mathbf{v}'/dt)$, with $d\mathbf{v}'/dt$ the observed acceleration in S'. This additional force $-m\,\mathbf{a}_0$ is known as a *fictitious force* and, provided that its effects are taken into account, the second law is applicable even in accelerated reference frames. Note, however, that a fictitious force is different from an ordinary one in that

it does not satisfy Newton's third law. Thus, if you are in an accelerated reference frame, you will experience the usual forces and exert back on their sources an equal and opposite force; but in addition you will experience the fictitious force $-m\mathbf{a}_0$ for which you will not be exerting an equal and opposite force! Nevertheless, these "fictitious forces" are very real and are just as capable of causing acceleration, deformation, or pain as ordinary forces.

Example 4-12 A man of mass $m = 100$ kg is in an elevator that starts from rest and accelerates upward with an acceleration 5.0 m/s^2. With what upward force does the elevator push the man?

Solution Let us first solve this problem in the x'-y' system fixed to the elevator (see Figure 4-19a). Here, the man is at rest and hence his acceleration $dv_y'/dt = 0$. Figure 4-19b shows the various forces acting on the man. These are: the unknown force $\mathbf{V}$ directed upward; the downward force $m\mathbf{g}$; and the fictitious force $-m\mathbf{a}_0$, which is also directed downward since $\mathbf{a}_0$ is upward. Substituting these forms into (4-26), we obtain

$$\mathbf{V} - m\mathbf{g} - m\mathbf{a}_0 = 0$$

since the man does not accelerate in this system. Solving for V and substituting the numerical values we find that

$$\begin{aligned} V = m(g + a_0) &= 100\text{ kg} \times (9.8 + 5.0)\text{ m/s}^2 \\ &= 1480\text{ N} \end{aligned}$$

Equivalently, we may solve the problem in the inertial x-y system. Here we find that

$$V - mg = m\frac{dv_y}{dt} = 100\text{ kg} \times 5.0\text{ m/s}^2$$

where the second equality follows since the man accelerates upward at 5.0 m/s^2 in this system. Solving for V we find the same value as above.

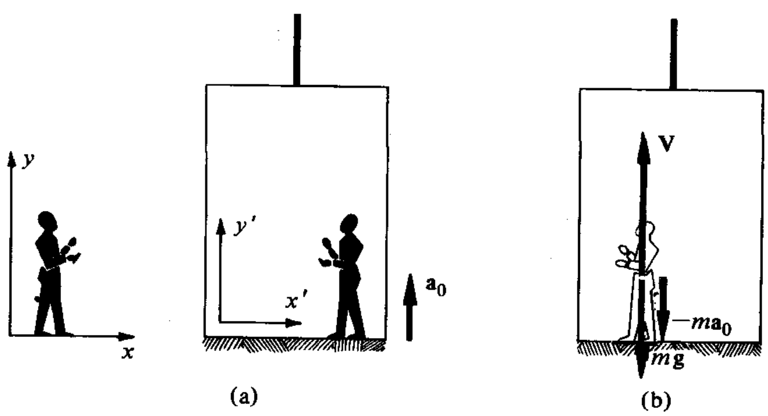


Figure 4-19

4-10 Other systems of units

The SI unit of force is the newton (N), which represents the force required to give an acceleration of 1 m/s^2 to a 1-kg body.

A second unit of force that is also frequently used is the *dyne* which is defined as the force that will give a 1-gram body an acceleration of 1 cm/s^2. The gram and the centimeter have been previously defined directly in terms of SI units. The relation between the dyne and the newton is thus easily calculated to be

$$1 \text{ dyne} = 10^{-5} \text{ N}$$

and thus 1 newton corresponds precisely to a force of 10^5 dynes. This system of units, which has the centimeter, the gram, and the second as its basic units, is known as the *CGS system*. In this system the acceleration of gravity g has the value $g = 980 \text{ cm/s}^2$.

A third system of units is the *English* (or *engineering*) *system*. Here, instead of defining the unit of force in terms of mass as we did above, one first defines a unit of force and then the unit of mass in terms of it. The unit of force in this system is the pound (lb) and is conventionally defined as the weight of a certain standard (body) at a fixed point on earth. A force of 1 lb is approximately 4.5 newton or, in more precise terms,

$$1 \text{ lb} = 4.4482 \text{ N}$$

The unit of mass in this system is the *slug*. This is defined to be the mass of a body that when subject to a force of 1 lb will undergo an acceleration of 1 ft/s^2. At a point on the earth where g has the value 32 ft/s^2, it thus follows that a body whose weight is 160 lb will have a mass m

$$m = \frac{w}{g} = \frac{160 \text{ lb}}{32 \text{ ft/s}^2} = 5.0 \text{ slugs}$$

Table 4-1 summarizes these three system of units and includes an approximate value for g in each system.

Table 4-1

	Force	*Mass*	*Acceleration*	g
SI	newton (N)	kg	m/s^2	9.8 m/s
CGS	dyne	g	cm/s^2	980 cm/s^2
English	pound (1b)	slug	ft/s^2	32 ft/s^2

4-11 Summary of important formulas

The basic results of this chapter are three physical laws, which are known as *Newton's laws*. These are:

Law 1: Every body continues in its state of rest or of uniform motion unless it is compelled by forces to change that state.

Law 2: If a body of mass m is subject to various forces and if $\mathbf{a}$ is its acceleration as observed in an inertial coordinate system, then

$$\mathbf{F} = m\mathbf{a} \tag{4-7}$$

where $\mathbf{F}$ is the vector sum of *all* the forces acting on the body.

Law 3: If body 1 exerts a force $\mathbf{F}_2$ on body 2 and the latter exerts a force $\mathbf{F}_1$ on the former, then regardless of what other forces may be acting on the two bodies these forces are equal and opposite:

$$\mathbf{F}_1 = -\mathbf{F}_2 \tag{4-10}$$

In addition, there is Newton's law of universal gravitation

$$F = \frac{Gm_1m_2}{r^2} \tag{4-15}$$

which relates the force of attraction (along the line joining them) between two particles of masses m_1 and m_2 and separated by a distance r.

The equation of motion of a particle with respect to a reference frame that has the acceleration $\mathbf{a}_0$ relative to an inertial coordinate system is

$$m\frac{d\mathbf{v}'}{dt} = \mathbf{F} - m\mathbf{a}_0 \tag{4-26}$$

where $d\mathbf{v}'/dt$ is the acceleration of the particle in this system, m is its mass, and $\mathbf{F}$ is the total force acting on it.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** inertial reference frame; **(b)** contact force; **(c)** additivity property of mass; and **(d)** weight.
2. You are given two bodies and asked to determine the ratio of their masses. Describe in operational terms how you can achieve this measurement.
3. Give the key experimental facts that underlie our definition of mass.
4. Give three examples of contact forces acting on bodies in your immediate environment. How many action-at-a-distance forces act on these same bodies?
5. Describe an experiment by means of which you could ascertain the force required to stretch a certain spring by 1 cm.
6. Why can we not define mass as the amount of matter present in a substance? Is there any experimental basis for this statement nevertheless? Explain.
7. A dictionary definition of inertia is "indisposition to motion." Explain in terms of this definition why mass can be thought of as a measure of inertia.
8. Contrast the two physical concepts of mass and weight. What are the two most important differences between them?
9. A student asks, "Since a locomotive pulls on a train with a force that is equal and opposite to the force with which the train pulls back on the locomotive, how can the train accelerate?" What response could you

give to help the student out of his difficulty?

10. Name the action-reaction pairs in the following: (a) a golf ball is struck by a club; (b) a raindrop falls on your umbrella; (c) an automobile accelerates along a highway; and (d) an airplane in flight changes direction by going in a circular path.
11. Even though the air around us is capable of exerting a contact force on any body in the atmosphere, we generally ignore this force. Why do you suppose this is so? Can you think of any circumstance for which this force cannot be neglected?
12. Under certain circumstances, astronauts describe themselves as being in a state of *weightlessness*. What does this mean? Why is it not possible for anyone to find himself ever in a corresponding state of *masslessness*?
13. Explain why no terrestrial laboratory can be a truly inertial frame.
14. If you were in an elevator falling freely at the acceleration of gravity g, you would find that if you suddenly shoved a body it would appear to continue to travel with its initial uniform velocity in a straight line (until it struck a wall). Does this elevator constitute an inertial reference frame? Explain.
15. Two astronauts are in a space vehicle that is circling the moon at an elevation of 100 km from its surface. They find that inside the vehicle a body at rest remains at rest and one in motion continues to move with uniform velocity in a straight line. Does this mean that the astronauts are in an inertial frame? Explain.
16. Describe briefly the extent to which the second law is definition and the extent to which it summarizes experimental facts.
17. It has been said that in the experiment in which Cavendish measured the gravitational constant G, he, in effect, "weighed the earth." Explain what is meant by this statement.
18. If $\mathbf{F}_1$ and $\mathbf{F}_2$ are the forces that two particles exert on each other, is it necessary in order to be consistent with the third law that $\mathbf{F}_1$ and $\mathbf{F}_2$ lie along the line joining the particles? Explain.
19. Explain why you tend to fall backward when the automobile you are in accelerates forward, and conversely you tend to fall forward when the automobile slows down.
20. Describe what happens to a vertically hanging curtain between the compartments in an airplane under the following conditions:
    (a) The plane accelerates on takeoff.
    (b) The plane cruises at a uniform speed.
    (c) The plane slows down in preparation for landing.
    Assume that the floor of the plane is horizontal under all circumstances.
21. A child sitting in an airplane holds in her hand a string to which is attached a floating helium-filled balloon. When the plane accelerates on takeoff, the balloon tends to move toward the *front* on the plane. Explain.
22. You are sitting in an airplane and hold a lighted match in your hand and note that when there is no acceleration, the flame rises vertically upward. When the plane is subsequently slowing down prior to landing, will the flame maintain this orientation or will it be inclined toward the front or back of the plane? Explain.
23. If you are a passenger on a ferris wheel, does the seat on which you sit exert a greater force on you when you are at the top of the wheel or at the bottom? Assume that the ferris wheel goes around with uniform speed.
24. A sphere of mass M hangs from a very light string, and an identical piece of string hangs from the bottom

of the sphere (see Figure 4-20). If the lower string is now placed under tension T_0, what is the tension in the upper string? As you gradually increase the tension T_0 in the lower string, why will the upper one snap first? Can you account for the fact that if you give the lower string a sudden jerk, then it—and not the upper one—will snap?

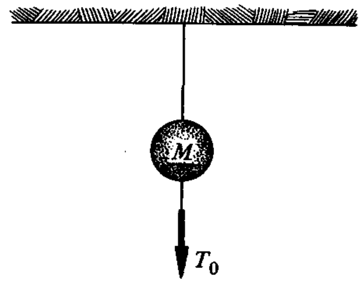


Figure 4-20

PROBLEMS

Note: For problems involving the usage of lunar or planetary data, see Appendix B.

1. A block of mass 3.0 kg lies on a horizontal and smooth surface and is subject to a constant horizontal force of 1.5 newtons.
   (a) What is the acceleration of the block?
   (b) If a second block of mass 2.0 kg is placed on top of the first and the same force is applied, what is now the acceleration of the system?
2. Two carts, one of mass 2.0 kg and the other of mass 3.0 kg, are connected by a very light spring and placed on a horizontal air track. If they are pulled apart and then released, then at an instant when the more massive one has an acceleration of 3.0 m/s^2 to the right:
   (a) What is the acceleration of the other cart?
   (b) What is the force acting on the 2.0 kg cart at this instant?
3. A 2.0 kg block is pushed along a rough horizontal surface by a force of 10 newtons. If the observed acceleration of the block is 3.0 m/s^2, what is the strength of the frictional force acting on the block?
4. An elevator has a mass of 7.0×10^3 kg and is subject to a certain upward force, transmitted by its supporting cable, so that it has an upward acceleration of 2.0 m/s^2. **(a)** With what force does the cable pull up on the elevator? **(b)** With what force does the cable pull up on the elevator when it has a downward acceleration of 2.0 m/s^2?
5. Consider a particle that travels along the trajectory

$$\mathbf{r} = 2t\mathbf{i} - 2t^2\mathbf{j}$$

   with t in seconds and $\mathbf{r}$ in meters. What force $\mathbf{F}$ is required to sustain this motion if the particle has a mass of 0.25 kg?
6. A body of mass 2.0 kg is subject to a force that acts in a fixed direction and whose magnitude in newtons is $F = 1.4t$, with t measured in seconds.
   (a) What is the acceleration of the body at time t?
   (b) Show that if the body starts from rest then at time t its velocity $v(t)$ in meters per second is

$$v(t) = 0.35t^2$$

   (c) How far will the body have traveled in 10 seconds?
7. An astronaut having a mass of 90 kg is in a spaceship, which on takeoff from the earth achieves an upward

acceleration of 30 m/s^2. (a) Calculate the upward force that the ship exerts on him. (b) If he were on a scale during this acceleration what would be its reading? By what factor is his weight greater than when he is at rest on earth?

8. A man of mass 80 kg drops from a vertical height of 20 meters into a mud puddle, so that he has a downward velocity of about 20 m/s just before he strikes the mud. (a) If he sinks a total distance of 10 cm into the mud, calculate his acceleration (assuming it to be constant) while he is sinking. (b) What upward force acts on the man while he is sinking into the mud?

9. An automobile of mass 1500 kg is traveling along a highway at 100 km/hr when the driver suddenly applies the brakes. Assume that the auto comes to rest after it has traveled an additional distance of 250 meters.
   (a) What is the (presumed constant) acceleration?
   (b) How long does it take for the auto to come to rest?
   (c) What force acts on the auto? At what point(s) on the automobile is this force applied?

10. A sphere of mass 3.0 kg is subject to two forces $\mathbf{F}_1$ and $\mathbf{F}_2$ as shown in Figure 4-21. Calculate the magnitude and direction of the acceleration of the sphere if $F_1 = 2.2$ newtons and $F_2 = 2$ newtons.

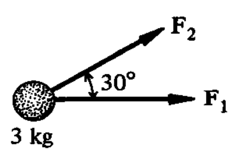


Figure 4-21

11. What force in addition to $\mathbf{F}_1$ and $\mathbf{F}_2$ must be applied to the body in Figure 4-6 so that:
    (a) It does not accelerate?
    (b) It has an acceleration of 2.0 m/s^2 along the negative x-axis? Assume that $F_1 = 2.0$ newtons and $F_2 = 1.0$ newton.

12. A proton of mass m enters the region between two capacitor plates, where it experiences a constant force $\mathbf{F}$. Initially the proton has a velocity $\mathbf{v}_0$.
    (a) Find its velocity at any time t if $\mathbf{v}_0$ is parallel to $\mathbf{F}$.
    (b) Find the distance the proton travels before coming to rest if $\mathbf{v}_0$ and $\mathbf{F}$ are antiparallel.

13. A bead of mass 5 grams is confined so that it moves with a uniform speed $v_0 = 10$ cm/s around a horizontal circle of radius $R = 20$ cm.
    (a) What is the acceleration of the bead?
    (b) Calculate the force required to sustain this motion.

14. Figure 4-22 shows a bead of mass m traveling with uniform speed v_0 around a thin, circular wire of radius R. Assume that the circle lies in a vertical plane so that the force of gravity acts vertically downward.

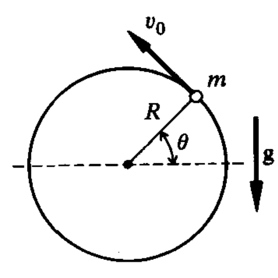


Figure 4-22

   (a) What is the centripetal acceleration?
   (b) Find the radial and tangential components of the force on the bead, due to the wire, at the instant when the radius to the bead makes an angle θ with the horizontal.

15. Assuming that the moon goes about the earth with a constant velocity in

a circular path of radius $R = 3.8 \times 10^5$ km once every 27.3 days, calculate **(a)** the moon's velocity in its orbit; **(b)** its acceleration; and **(c)** the force required to sustain this circular motion.

16. Calculate the gravitational force of attraction between the earth and the moon and compare this with the observed motion of the moon around the earth. (See Problem 15.)

17. Figure 4-23 shows a small object of mass 100 grams suspended by a massless string of length 20 cm and traveling with uniform speed along a horizontal circular orbit of radius 10 cm.

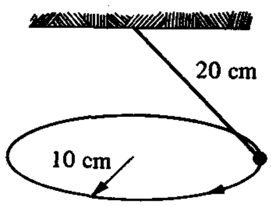


Figure 4-23

(a) Using the fact that there is no acceleration along the vertical, calculate the tension in the string.

(b) Calculate the speed of the particle as it traverses its orbit.

18. According to classical ideas, a hydrogen atom consists of a proton with an electron ($m = 9.1 \times 10^{-31}$ kg) orbiting about it with a uniform velocity of 2.2×10^6 m/s in a circle of radius 5.3×10^{-11} meter.

(a) What is the acceleration of the electron?

(b) What force is required to sustain this motion?

(c) Compare your answer to (b) with the gravitational force between the proton and the electron.

19. A man of mass 100 kg lives at the equator.

(a) To what centripetal acceleration is he subjected because of the rotation of the earth about its axis?

(b) What force is required to produce this acceleration?

(c) Is he lighter or heavier than he would be if the earth were not rotating? By how much?

20. Repeat Problem 19, but this time assume the man to live at a latitude λ. (*Note:* If you live at a latitude λ, the radius of your orbit is $R \cos \lambda$, with R the radius of the earth.)

21. Using the data in Appendix B, calculate the gravitational force of attraction between the earth and the sun. Compare your answer with that of Example 4-5.

22. Using the data in Appendix B, calculate the gravitational force which the overhead sun exerts on a person of mass 90 kg. Compare this with his weight.

23. An astronaut whose weight on earth is 700 newtons lands on the planet Venus. He steps on a scale there and notes that, after making corrections for the equipment he wears, he now weighs 600 newtons. Using these data and the fact that the diameter of Venus is about the same as that of the earth, calculate the mass of the planet Venus. Compare with the value in Appendix B.

24. Calculate the ratio of your weight on the planet Mars to that on earth.

25. At closest approach, the distance between the earth and Mars is about 8.0×10^7 km. Calculate the force between these two planets at this separation and compare this with the force that the sun exerts on the earth.

26. If w_0 is the weight of a man on the surface of the earth, what is his weight $w(r)$ when he is at a distance r above its surface? At what height above the earth will his weight be $\frac{1}{2}w_0$?

27. Consider again the physical situation described in Figure 4-14a, but suppose now that the same horizontal force **F** pushes B to the left. What is the magnitude of the force that the blocks exert on each other this time? Assume $F = 3$ newtons.
28. Three blocks, A, B, and C, of respective masses 2 kg, 3 kg, and 4 kg are lined up in contact on a smooth, horizontal surface (see Figure 4-24).

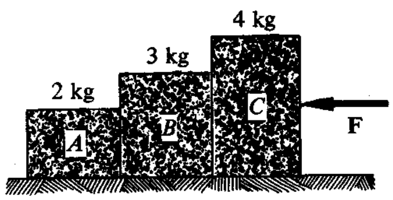


Figure 4-24

   (a) Find the value of the constant force F applied to C so that the system will have an acceleration of 2 m/s^2.
   (b) Calculate the force that B exerts on A.
   (c) Calculate the force that A exerts on B.
   (d) Calculate the force that B exerts on C.
29. Repeat Problem 28, but suppose now that the force **F** in Figure 4-24 acts on A and pushes the system of blocks to the right.
30. Suppose that the block in Figure 4-10 has a mass M and the rope has a mass m. If the block lies on a smooth surface and the rope is pulled by a constant force F, calculate the acceleration of the system and the force with which the rope pulls on the block.
31. Three identical blocks each of mass 300 grams are connected by massless strings, as shown in Figure 4-25. Assume that they lie on a smooth, horizontal surface and are observed to have an acceleration of 2.0 m/s^2 under the action of a force **F** acting to the right. (a) Find the value for F. (b) Calculate the tension in each of the connecting strings.

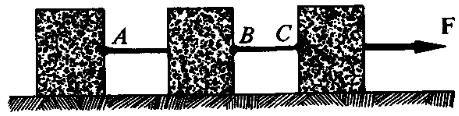


Figure 4-25

32. Consider the physical situation in Problem 31, but suppose that each connecting rope has a mass of 100 grams.
   (a) What force is now required to produce the acceleration of 2.0 m/s^2?
   (b) Calculate the forces exerted by the ropes on the blocks at the points A, B, and C in Figure 4-25.
33. A block of mass 2 kg lies on a smooth, horizontal surface and is connected by two massless strings which hang over two smooth pegs to two weights of respective masses 1 kg and 3 kg (see Figure 4-26).

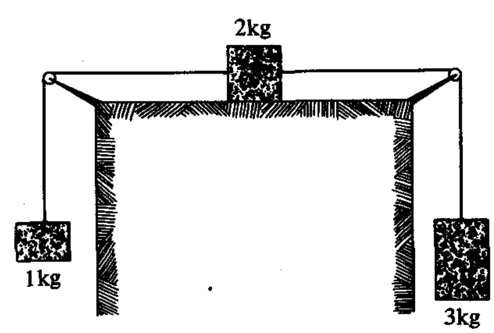


Figure 4-26

   (a) Introduce appropriate symbols for the tension in each string and apply the second law to each mass.
   (b) Show that the acceleration of the system is $g/3 = 3.3$ m/s^2 clockwise.
   (c) Calculate the tension in each string.
34. Calculate the acceleration of an Atwood's machine if one of the two

weights is twice as heavy as the other. If the heavier of the bodies has a mass of 5.0 kg, what is the tension in the supporting rope?

35. An Atwood's machine, as shown in Figure 4-27, has attached to it two buckets full of sand, having respective total masses of 1.5 kg and 1.0 kg. A 0.5-kg stone is attached to the string supporting the less massive bucket, at a distance of 10 cm above the sand in this bucket.

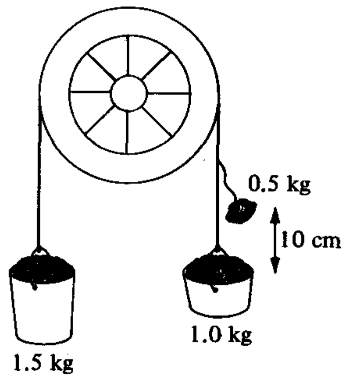


Figure 4-27

(a) What is the initial acceleration of the system?
(b) What is the tension in the various parts of the string?

*36. For the physical situation described in Problem 35, suppose that at $t = 0$ the string supporting the 0.5-kg stone is cut, so that it falls down into the sand in the lighter bucket.
(a) What is the downward acceleration of the stone just before it strikes the sand?
(b) What is the upward acceleration of the 1-kg bucket at the instant the stone lands?
(c) At what time t does the stone fall into the bucket?
(d) Describe what happens to the system after the stone has landed.

37. In the Atwood's machine shown in Figure 4-16a, suppose that $m_1 = 100$ grams and that m_2 consists of two 100-gram blocks, one of which hangs below the other by means of a massless string. Calculate the tension in this connecting string and the acceleration of the system.

*38. Figure 4-28 shows an Atwood's machine where the body on one side is itself an Atwood machine. Let the ropes and the supporting wheels be massless and assume that all contacts are smooth.

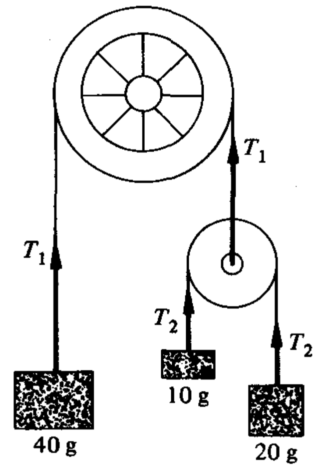


Figure 4-28

(a) Why must the tensions T_1 and T_2 satisfy $T_1 = 2T_2$?
(b) If a_0 is the downward acceleration of the 40-gram block, write down a relation between T_1 and a_0.
(c) Write down the equation of motion of the 10- and 20-gram blocks. (*Hint:* If $(a - a_0)$ is the downward acceleration of the 20-gram block, what must be the upward acceleration of the 10-gram block?)
(d) Calculate values for T_1, T_2, a_0, and a.

†39. An astronaut who weighs 720 newtons is in a spaceship that during takeoff from the earth has an upward acceleration of 30 m/s^2.
(a) Calculate the upward force on the astronaut by use of a system fixed to the earth.

(b) Repeat (a), but this time by use of a coordinate system fixed to the ship.

(c) How would your answers to (a) and (b) differ if this acceleration took place in outer space?

†40. Consider the system in Figure 4-26, but suppose that now it has an upward acceleration of $300\,cm/s^2$. Find the tension in each string and the acceleration of each mass.

†41. A weight of mass 500 grams is suspended from a spring scale attached to the ceiling of an elevator. Give the scale reading under the following conditions: **(a)** The elevator goes up with uniform velocity. **(b)** The elevator accelerates upward at $0.3\,m/s^2$. **(c)** The elevator accelerates downward at $0.2\,m/s^2$.

†42. If the Atwood machine in Figure 4-16 had an upward acceleration $\mathbf{a}_0$, what would be the tension in the rope and the acceleration of each mass?

5 Applications of Newton's laws

...It is by the solution of problems that the strength of the investigator is hardened; he finds new methods and new outlooks and gains a wider and freer horizon.

DAVID HILBERT

5-1 Introduction

In Chapter 4 we described Newton's three laws of motion and illustrated their usage by reference to several physical situations. We continue these applications in this and the next chapter with the main emphasis now on the case of a single particle moving under the influence of a fixed force. The associated problem of more than one particle will be considered in Chapters 11 and 12.

Before we go into detail on the dynamics of a single particle, however, it is of some interest to consider first some problems in the field of *statics*. These deal with the case of a body that may be subject to the simultaneous action of several forces but nevertheless, does not accelerate. According to Newton's second law the vector sum of the forces acting on such a body must vanish, and this requirement of the vanishing of the total force acting on a body is then the basic predictive tool of statics.

The remainder of the chapter is then devoted to the more general problem of predicting the motion of a particle when it is subject to a given

nonvanishing force **F**. Our main task here is that of solving the Newtonian *equation of motion*

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} \tag{5-1}$$

for a variety of force functions in order to obtain the associated position functions $\mathbf{r}(t)$ of the particle at time t. The problem of calculating the elliptical path of the earth about the sun by use of the gravitational force in (4-15) is an important problem of this type. Generally speaking, given an arbitrary force function **F**, it is *not* possible to solve (5-1) explicitly for the position of the particle. However, for certain very special choices for **F**, the trajectory of the particle can be explicitly obtained, and in this chapter and the next we shall consider a variety of these cases.

5-2 Statics of a particle

Most objects in your immediate environment, such as the chair on which you sit, the book you hold in your hand, or the building in which you reside, do not accelerate under normal circumstances. It follows from Newton's law that the net or total force acting on each of these objects must vanish. *Statics* is that branch of mechanics whose concern is the study of nonaccelerating bodies, and the purpose of this section is to describe briefly some problems in this field.

The basic principle of statics may be described by the following statement:

If a body is not accelerating, so that it is either at rest or moving with uniform velocity in a straight line, then the vector sum of all forces acting on it must vanish.

A body that is not accelerating is said to be in a state of *static equilibrium* or in *equilibrium*. Thus the basic principle of statics may be described equivalently by saying that a necessary condition for a body to be in static equilibrium is that the vector sum of all forces acting on it vanishes.

Figure 5-1a shows a body that is maintained in a state of equilibrium under the simultaneous action of the three forces $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$. According to the principle of statics, it follows that

$$\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = 0 \tag{5-2}$$

and therefore a knowledge of any two of these three forces enables us to deduce the third. In other words, we have here a means for measuring force without resorting to the indirect dynamical procedures of Chapter 4. When you step on a scale, for example, you make explicit use of this principle to determine your weight. Since the resultant of the three vectors $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$

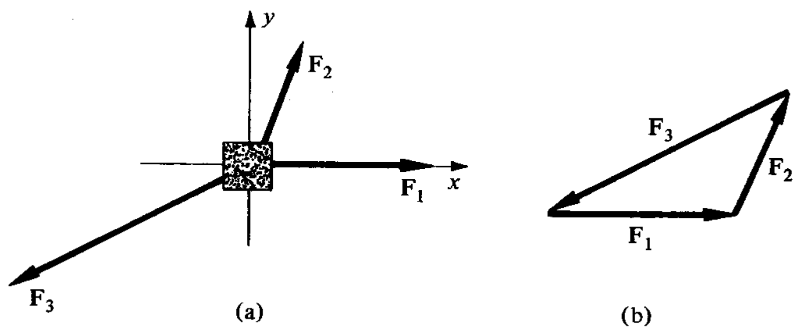


Figure 5-1

vanishes, it follows, as shown in Figure 5-1b, that they form the sides of a closed triangle. More generally, making use of the graphical method for adding vectors, as in Figure 3-9, it follows that if a body is in equilibrium under the combined action of n forces, $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_n$, then the arrows representing these forces may be arranged so that they form a closed, n-sided polygon.

From a practical point of view, the kind of problems that can be solved in statics may be seen by reference to the situation in Figure 5-1. If we set up a Cartesian coordinate system in the plane of the three vectors, then (5-2) may be expressed in its component form

$$\begin{aligned} F_{1x} + F_{2x} + F_{3x} &= 0 \\ F_{1y} + F_{2y} + F_{3y} &= 0 \end{aligned} \tag{5-3}$$

with F_{1x} the component of $\mathbf{F}_1$ along the x-axis and similarly for the remaining components. The relations in (5-3) represent two relations among a total of the six components of $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$. Hence if any four of these are known, the remaining two may be determined.

Example 5-1 A 2.0-kg block is suspended by a massless string, which is maintained at an angle of 30° with respect to the vertical by a horizontal force $\mathbf{F}$. Find the value for F and the tension T in the string.

Solution Figure 5-2a shows the physical situation, and in part (b) the block has been isolated and the forces acting on it, including its own weight $m\mathbf{g}$, have been indicated

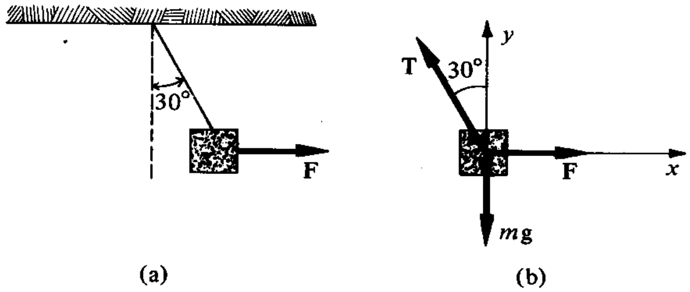


Figure 5-2

by arrows. The condition for equilibrium is

$$\mathbf{F}+\mathbf{T}+m\mathbf{g}=0$$

Setting up a coordinate system as shown in the figure we find the components of this relation along the x- and y-axes to be

$$F - T\sin 30° = 0$$
$$T\cos 30° - mg = 0$$

Since $\sin 30° = 0.5$ and $\cos 30° = 0.87$, the second equation leads to

$$T=\frac{mg}{\cos 30°}=\frac{2.0\ \text{kg}\times 9.8\ \text{m/s}^2}{0.87}=23\ \text{N}$$

and, substituting into the first equation, we obtain to two-place accuracy

$$F = 0.5\,T = 12\ \text{N}$$

Thus we are able to deduce a value for F and for the tension T in the string given only the weight mg of the block.

Example 5-2 A student of mass $m = 80$ kg hangs by his hands from a horizontal gymnastics bar (Figure 5-3a). If each of his arms makes an angle $\theta = 30°$ with the vertical, with what force must he pull on the bar with each of his hands?

Solution Figure 5-3b shows the student replaced by a point particle of mass m and the various forces acting on him, which include the force of gravity $m\mathbf{g}$, acting vertically down, and the forces $\mathbf{F}_1$ and $\mathbf{F}_2$, which the horizontal rod exerts on each of his hands. (By the third law, $-\mathbf{F}_1$ and $-\mathbf{F}_2$ are the forces he exerts on the rod.) Since he does not accelerate,

$$\mathbf{F}_1+\mathbf{F}_2+m\mathbf{g}=0$$

Taking components, we find that

$$F_1\cos\theta + F_2\cos\theta - mg = 0$$
$$F_1\sin\theta - F_2\sin\theta = 0$$

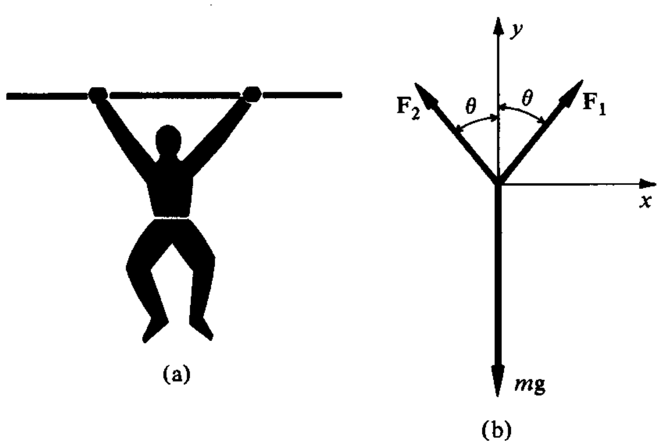


Figure 5-3

The second of these implies that $F_1 = F_2$, a fact that could have been anticipated by reason of symmetry. Substituting this result into the first equation, we obtain

$$F_1 = F_2 = \frac{mg}{2\cos\theta}$$

which, by use of the given values of m and θ, leads to the desired result

$$F_1 = \frac{80\text{ kg} \times 9.8\text{ m/s}^2}{2 \times 0.87} = 450\text{ N}$$

since $\cos 30° = 0.87$.

Example 5-3 A wooden block of mass m is at rest on an inclined plane of elevation angle α, under the action of a force **F** which acts upward and parallel to the plane (see Figure 5-4). Calculate the magnitude and direction of the force which the plane exerts on the block.

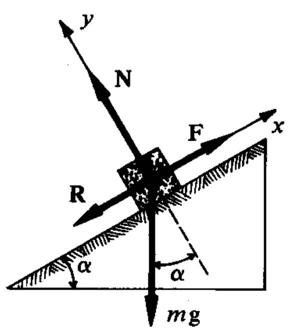


Figure 5-4

Solution Let N be the normal component and R the corresponding tangential component of this unknown force and set up a coordinate system, as shown in the figure, with axes parallel and perpendicular to the plane. On taking the components of the equilibrium condition

$$\mathbf{N} + \mathbf{R} + \mathbf{F} + m\mathbf{g} = 0$$

along these two axes, we find that along the x-axis

$$F - R - mg\sin\alpha = 0$$

and, correspondingly, that along the y-axis

$$N - mg\cos\alpha = 0$$

The solution of these for the unknown components N and R is

$$N = mg\cos\alpha \qquad R = F - mg\sin\alpha$$

Hence, according to (3-8) and (3-9), the magnitude of the force that the block exerts on the plane is

$$\sqrt{(mg\cos\alpha)^2 + (F - mg\sin\alpha)^2}$$

and the angle θ it makes with the positive x-axis is given by

$$\tan\theta = -\frac{N}{R} = -\frac{mg\cos\alpha}{F - mg\sin\alpha}$$

where the negative sign is due to our choice of having R point down the plane. Note that if $F < mg\sin\alpha$, then $R < 0$, and this means that contrary to the figure, **R** points *up* the plane.

5-3 Friction

In our studies of Newton's laws in Chapter 4 we made use of certain idealized experiments for which the effects of friction could be neglected. Such idealized conditions do not generally prevail in practice, and for laboratory experiments involving macroscopic bodies the forces of friction normally play a significant role. Furthermore, a world without friction is one in which we could not walk or hold a pencil, and, so forth, and as such, is inconceivable. Fortunately, although frictional forces are often an unpleasant nuisance in the laboratory, they can be dealt with very simply in terms of a certain phenomenological parameter, called the *coefficient of friction*, and in this section we shall describe this quantity.

Figure 5-5a shows a block of mass m at rest on a rough, horizontal surface. Since the force of gravity mg acts vertically downward and there is no acceleration, it follows that the upward force **N** that the surface exerts on the block has the magnitude mg. Let us now take this same block and, as shown in Figure 5-5b, apply to it a certain horizontal force **F**, which is sufficiently small that the block does not move. Under this circumstance, in addition to the normal force **N**, the surface must also exert on the block a horizontal force **f**, which in accordance with the principle of statics must be equal and opposite to **F**. We call this force **f** which the surface exerts on the block the *force of friction.*

Suppose that we now steadily increase the strength of the horizontal force **F** in Figure 5-5b. The magnitude of the frictional force **f** *will also increase, and in such a way that at each stage it is equal and opposite to* **F**. Ultimately, the frictional force f will achieve a certain maximum value f_0, at which the block

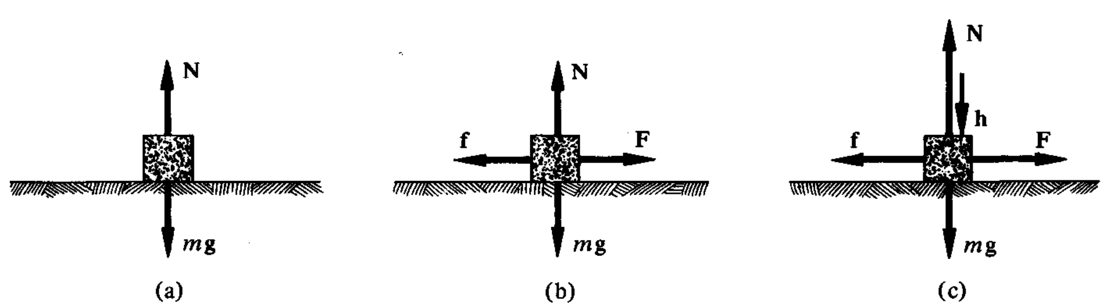


Figure 5-5

first begins to slip. We define the *coefficient of static friction* μ_s to be the ratio of this maximum friction force f_0 and the magnitude of the normal force N. Expressed in mathematical terms, we have

$$f_0 = \mu_s N \tag{5-4}$$

Experiment shows that, within limits, μ_s is independent both of the size of the surface areas in contact with each other and of the normal force N. If, as illustrated in Figure 5-5c, for example, the normal force N is increased by pressing down on the block with a certain force **h**, we find the same value for μ_s as before. That is, if N is increased by pressing down on the block, then the maximum frictional force f_0 will also increase but in a way so that the ratio f_0/N ($\equiv \mu_s$) remains the same. Note that as long as the frictional force **f** has a magnitude less than $f_0 \equiv \mu_s N$, there is no motion of the block. Only when f achieves its maximum value f_0 does the block begin to slip.

Values of μ_s for a few surfaces are listed in Table 5-1. Note from this table that μ_s depends generally on the state of the two surfaces in contact. From a practical point of view, this means that it is generally advisable to measure μ_s directly if its value is required. Example 5-4 suggests a simple and practical way for measuring μ_s.

Table 5-1 Coefficient of static friction μ_s*

Materials	*Condition*	μ_s
Ice on ice	Clean	0.05–0.15
Ski wax on snow	Wet	0.1
Ski wax on snow	Dry	0.04
Glass on glass	Clean	0.9–1.0
Glass on glass	With alcohol	0.1
Steel on steel	Clean	0.6
Steel on steel	With castor oil	0.1
Graphite on steel	Clean	0.1

*Adapted from *Handbook of Chemistry and Physics*, 50th ed. Cleveland: Chemical Rubber Publishing Co., 1970.

In order for the definition of μ_s in (5-4) to be useful in a practical way, it is necessary that it be a constant for any two given surfaces under specified conditions. Consistent with this expectation, experiment shows that μ_s is indeed independent of the area of the surfaces in contact and, within limits, also of the strength of the normal force N. Of course, if N becomes so large that the surface becomes deformed, then the simple relation in (5-4) ceases to be valid.

An additional source of variation in the frictional force occurs if the two surfaces are in motion. Once the magnitude of the force **F** in Figure 5-5 achieves the strength f_0, and the block begins to slide, experiment shows that

there is a sudden *decrease* in the magnitude of the frictional force f. That is, once the surfaces are in relative motion, we find that the associated frictional force f_k attains a magnitude *less than* f_0. It is, however, still related to the normal force N by

$$f_k = \mu_k N \tag{5-5}$$

with μ_k a constant known as the *coefficient of kinetic* or *sliding friction.* If, for example, a block of mass m is pulled along a horizontal surface by a force **F** (see Figure 5-6), then

$$F - \mu_k N = ma_x$$
$$N - mg = 0$$

where a_x is the acceleration along the horizontal and where the second relation follows since there is no motion along the vertical. In general, given any two surfaces, we find that the coefficient of kinetic friction μ_k is 20 to 50 percent less than is the value of μ_s for the same two surfaces. And just as for μ_s, we find that μ_k depends only on the nature of the two surfaces and is not dependent on their size nor on the strength of the normal force N acting between them. Table 5-2 lists values for μ_k for a sampling of surfaces.

A very important feature of frictional forces that is worth emphasizing is that the force of friction is always directed opposite to the direction of actual or of potential motion. If, in Figure 5-7a, a block slides to the left over a rough, horizontal surface, then the frictional force is directed to the right and has magnitude $\mu_k N$. Similarly, if the same block slides to the right (Figure 5-7b),

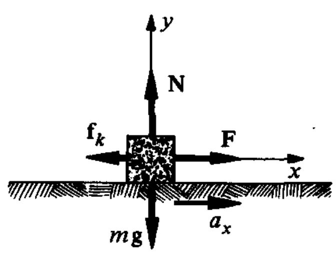


Figure 5-6

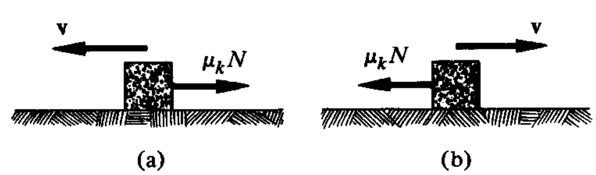


Figure 5-7

Table 5-2 Coefficient of kinetic friction μ_k*

Materials	*Relative speed*	μ_k
Waxed hickory–dry snow	0.1 m/s	0.04
Waxed hickory–wet snow	0.1 m/s	0.14
Ice–ice (clean)	4 m/s	0.02
Ebonite–ice (clean)	4 m/s	0.02
Natural rubber–ground glass	2 m/s	1.1

*Adapted from *Handbook of Chemistry and Physics,* 50th ed. Cleveland: Chemical Rubber Publishing Co., 1970.

then the frictional force of strength $\mu_k N$ is directed to the left. Thus the force of friction depends on velocity, and in this sense it is similar to certain other velocity-dependent forces that occur in nature.

Friction is a purely macroscopic phenomena, and exists only between two macroscopic surfaces in contact. There is no force of friction at the microscopic level.

Example 5-4 A block of mass m rests on a rough, inclined plane. Show that as the angle of inclination of the plane is increased, if the block first starts to slide down when its angle is ϵ, then

$$\tan \epsilon = \mu_s$$

where μ_s is the coefficient of static friction.

Solution Figure 5-8 shows the situation at the instant when the block is just about to slide down the plane. Since the block is still not accelerating, it follows that the total force **R** that the plane exerts on the block must be equal and opposite to its weight $m\mathbf{g}$.

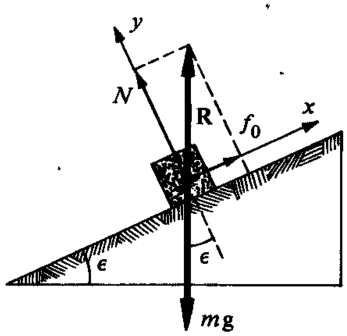


Figure 5-8

Further, the component of **R** along the x-axis must have the value f_0 in (5-4), whereas the normal force N is the component of **R** along the y-axis. Thus

$$\mu_s = \frac{f_0}{N} = \frac{R_x}{R_y} = \frac{R \sin \epsilon}{R \cos \epsilon} = \tan \epsilon$$

and this is the sought-for relation.

5-4 Motion under the action of frictional forces

In this section we apply the ideas of the preceding one to physical situations involving bodies undergoing acceleration.

Example 5-5 A 5-kg block is given a shove so that it starts to slide to the right in Figure 5-9 across a rough, horizontal surface at an initial speed of $v_0 = 4.0$ m/s. If the coefficient of sliding friction $\mu_k = 0.2$, what distance will it travel before coming to rest?

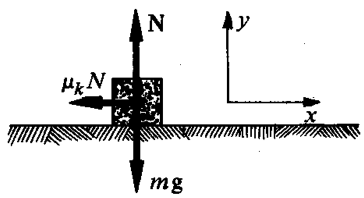


Figure 5-9

Solution Making use of the coordinate system shown in the figure, Newton's second law, and the fact there is no motion along the y-axis, we find that

$$N - mg = 0$$
$$-\mu_k N = ma_x$$

where the minus sign in the second relation is due to the fact that the frictional force is oriented along the negative x-axis. Eliminating the normal force N between these two relations and canceling the common factor m, we find that

$$a_x = -\mu_k g$$

that is, the block travels at the constant acceleration $-g\mu_k$. Substitution into the constant-acceleration formulas, (2-10) and (2-11), leads to

$$v_x = -\mu_k g t + v_0$$
$$x = -\frac{1}{2}\mu_k g t^2 + v_0 t$$

where v_0 is the initial velocity of the block when it is at the origin. If τ is the time it takes the block to come to rest, then according to the first relation, $\tau = v_0/g\mu_k$. Substitution into the second relation shows that the distance D the block travels before coming to rest is

$$\begin{aligned} D = x(\tau) &= -\frac{1}{2}\mu_k g \tau^2 + v_0 \tau \\ &= -\frac{1}{2}\mu_k g\left(\frac{v_0}{\mu_k g}\right)^2 + v_0\frac{v_0}{\mu_k g} \\ &= \frac{1}{2}\frac{v_0^2}{\mu_k g} = \frac{1}{2}\frac{(4.0\ \text{m/s})^2}{0.2 \times 9.8\ \text{m/s}^2} \\ &= 4.1\ \text{m} \end{aligned}$$

Example 5-6 A child starts out at rest at the top of a 45° slide having a total length of 7.0 meters. If the coefficient of friction between the child and the slide is 0.15, calculate:

(a) The child's acceleration down the slide.
(b) The child's velocity at the bottom.

Solution Figure 5-10a shows the physical situation, and in Figure 5-10b the forces acting on the child are represented by arrows.

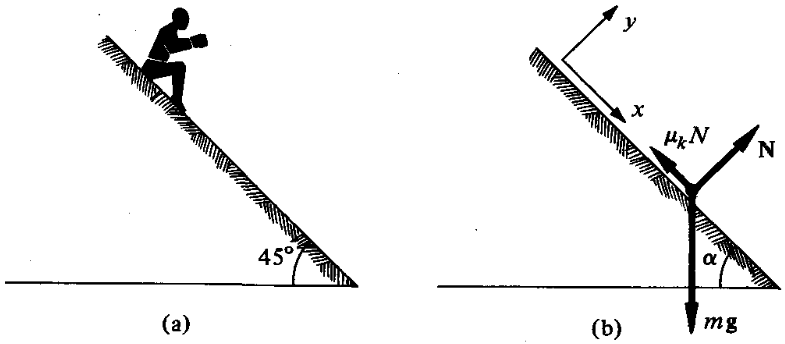


Figure 5-10

(a) As in the previous example, for the given choice of coordinate axes, with the x-axis parallel to and directed down the plane, from Newton's second law we have

$$mg \sin \alpha - \mu_k N = ma_x$$
$$N - mg \cos \alpha = 0$$

where a_x is the child's acceleration down the plane. If we eliminate N, the unknown mass m drops out with the result

$$\begin{aligned} a_x &= g(\sin \alpha - \mu_k \cos \alpha) \\ &= (9.8 \text{ m/s}^2) \times (\sin 45° - 0.15 \cos 45°) \\ &= 5.9 \text{ m/s}^2 \end{aligned}$$

(b) Since the child starts out at rest, from (2-18) it follows that

$$v_x = \sqrt{2xa_x}$$

where g has been replaced by the constant a_x, and x is the distance down the plane. Hence, at the bottom of the slide, $x = 7.0$ meters, and the child's velocity v is

$$\begin{aligned} v &= \sqrt{2 \times 7.0 \text{ m} \times 5.9 \text{ m/s}^2} \\ &= 9.1 \text{ m/s} \end{aligned}$$

5-5 The uniform force field

One of the very simplest types of forces that a particle can experience is a force **F** that is constant and thus independent of its position **r**, its velocity **v**, and time t. When the net or total force acting on a particle is of this nature, the particle is said to undergo motion in a *uniform force field.* There are two physical situations which are of particular interest in this connection. One of these deals with the motion of a particle in the uniform gravitational field of the earth—or of any other planet—near its surface, and the second deals with the motion of an electrically charged particle in the uniform electric field that

exists between the parallel plates of a capacitor. The predictions made by use of Newton's second law for these two cases are mathematically the same, so here we will consider only the case of the uniform gravitational field.

Figure 5-11 shows a particle of mass m at a distance $r = R + Z$ from the center C of the earth. According to the law of universal gravitation in (4-15), the particle experiences a radially downward force $\mathbf{F}$ of magnitude

$$F = \frac{GM_E m}{(R+Z)^2} \tag{5-6}$$

with G the gravitational constant, M_E the mass of the earth, and R the radius of the earth. For trajectories that are within, say, a kilometer or two of the earth's surface, to a very high degree of accuracy (of the order of 0.02 percent), we may neglect the particle's height Z in comparison with the radius R of the earth. For these cases, the force F in (5-6) becomes

$$\begin{aligned} F &= \frac{GM_E m}{R^2} \\ &= mg \end{aligned} \tag{5-7}$$

where the last equality follows from (4-19). In other words, for particle trajectories within a kilometer or two of the earth's surface, the gravitational force on the particle can be approximated by its weight mg and, since this is constant, the particle can be thought of as being in a uniform field throughout its motion.

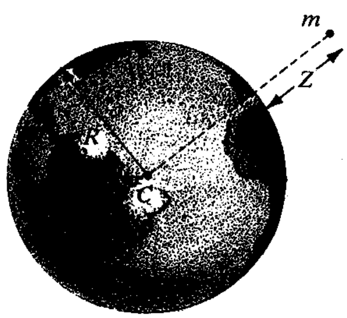


Figure 5-11

To obtain the equation of motion for the particle in the uniform gravitational field, let us substitute the force mg in (5-7) into Newton's second law. Canceling[1] the factor m we obtain

$$\frac{d\mathbf{v}}{dt} = \mathbf{g} \tag{5-8}$$

and this is the basis of the various projectile motion formulas in Section 3-11.

In our subsequent studies of electrostatics, we shall find that if a particle of

[1]See, in this connection, the discussion of the equivalence between inertial and gravitational mass in Section 5-9.

charge q travels through an electric field $\mathbf{E}$, then the force $\mathbf{F}$ on it is $q\mathbf{E}$. Accordingly, if the field $\mathbf{E}$ is uniform, then the force $q\mathbf{E}$ is constant, and thus by the simple replacement

$$m\mathbf{g} \rightarrow q\mathbf{E}$$

all results derived for the uniform gravitational field in Section 3-11 may be used for the electric field as well.

5-6 One-dimensional motion in the inverse-square field

As noted in Section 5-5, only if the trajectory of a body is confined to a region near the earth's surface can that body be said to be moving in a uniform force field. In this section we shall treat the more general case of a body in the gravitational field of a planet but not confined to a region near its surface. This means that the force acting on the body cannot be assumed to be constant; rather, in accordance with (4-15), this force varies depending on its location relative to the planet's center. For reasons of conceptual and mathematical simplicity, only the special case for which the motion is confined to a single line along a radius vector will be discussed here. The more general case, which deals with the motions of the planets, the satellites of planets, and so on, will be treated in Chapter 10.

Consider, in Figure 5-12, a "small" body of mass m, which is acted on by the gravitational attraction of a planet of mass M ($\gg m$), and is confined to motion along a single direction, the y-axis in the figure. Let us select an origin at the center of the planet and suppose that initially the particle is at a distance a from the origin and moving away from it at a velocity v_0. If $y = y(t)$ is the distance of the particle from the origin at time t, then according to the law of universal gravitation, the force F acting on it at this instant is

$$F = -\frac{GMm}{y^2} \tag{5-9}$$

where the minus sign reflects the fact that $\mathbf{F}$ is directed along the negative sense of the y-axis. Substituting this formula into Newton's second law and

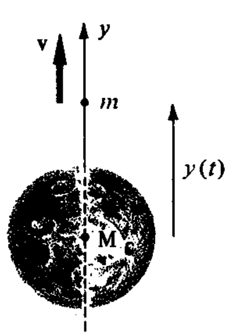


Figure 5-12

canceling the common factor m, we obtain the equation of motion

$$\frac{dv}{dt} = -\frac{GM}{y^2} \tag{5-10}$$

where $v = dy/dt$ is the velocity of the particle at time t. Note that here the acceleration of the particle is not given directly as an explicit function of time, but only indirectly through the implicit dependency of y on t. It follows that (5-10) cannot be integrated as directly as the constant-acceleration formula. However, it can be used to obtain a relation between v and y. Since many questions of physical interest can be answered by the resulting formula, let us derive this relation.

To this end, let us multiply both sides of (5-10) by the quantity $v\,dt$. The result may be written

$$v\,dv = -\frac{GM}{y^2}\,dy$$

where on the left side the identity $v\,dt(dv/dt) = v\,dv$ has been used, and on the right side the fact that $v\,dt = dy$. Integrating, we have

$$\int v\,dv = -GM\int \frac{dy}{y^2}$$

and since

$$v = \frac{d(v^2/2)}{dv} \qquad \text{and} \qquad -\frac{1}{y^2} = \frac{d}{dy}\left(\frac{1}{y}\right)$$

this may be cast into the form

$$\frac{1}{2}v^2 = \frac{GM}{y} + c$$

with c an integration constant. Finally, since at $t = 0$, $y = a$, $v = v_0$, we find that c has the value

$$c = \frac{1}{2}v_0^2 - \frac{GM}{a}$$

and thus a first integral of (5-10) is

$$\frac{1}{2}v^2 - \frac{GM}{y} = \frac{1}{2}v_0^2 - \frac{GM}{a} \tag{5-11}$$

Although it is possible to integrate (5-11) once more to obtain y as an explicit function of time, the resulting formula is somewhat complex and will not be described here.

Figure 5-13 shows a plot of v as a function of y as predicted by (5-11). The curve, labeled A, corresponds to the initial condition $y = a$, $v_0 = 0$. In this case the body starts out at rest at $y = a$, and descends towards the planet with ever increasing speed. Curve B depicts the case $v_0 > 0$. Here the particle starts out

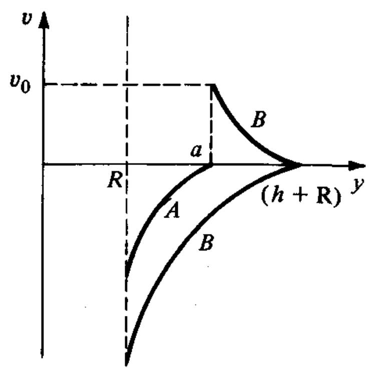


Figure 5-13

at the point $y = a$, moves away from the planet with decreasing speed until it comes to rest at a certain height h above its surface. Then it behaves as would a particle starting out at rest at a height h above the planet's surface. All motion ceases with impact on the planetary surface at $y = R$.

Example 5-7 Consider the situation in Figure 5-12, with the particle originally at the point $y = a$ and moving away from the planet at the velocity v_0. Calculate:
(a) The maximum distance h from the planet that the particle can reach.
(b) Its speed when it is at a distance $a/2$ from the origin.

Solution Physically, what is taking place is the following. Originally, the particle moves away from the planet, but because of the attractive force, this speed of attempted escape continually decreases, until a point is reached at which the particle comes to rest. After this it turns around and falls back towards the planet with ever-increasing speed ($v < 0$).

(a) The maximum separation distance of the particle, h, is achieved at the point where its velocity v is zero. Setting $v = 0$, $y = h$ in (5-11), we find that

$$-\frac{GM}{h} = \frac{1}{2} v_0^2 - \frac{GM}{a}$$

whence

$$h = a\frac{1}{1-(av_0^2/2GM)}$$

Note that since M is very large, h is always greater than a, as it must be.

(b) The velocity v_1 of the particle when it is at the distance $a/2$ from the planet's center is obtained by setting $y = a/2$ and $v = v_1$ in (5-11). This leads to

$$\frac{1}{2} v_1^2 - \frac{GM}{a/2} = \frac{1}{2} v_0^2 - \frac{GM}{a}$$

or, in other words,

$$v_1 = -\sqrt{v_0^2 + \frac{2GM}{a}}$$

where the choice of sign in front of the radical is due to the fact that when the particle is at the point $a/2$ it is moving toward the origin.

Example 5-8 An asteroid of mass 10^7 kg is at rest at a distance of 10^8 meters from the center of the earth. Calculate the velocity with which it would strike the earth's surface if the earth had no atmosphere.

Solution In terms of the parameters in (5-11) we are given the values

$$v_0 = 0 \qquad a = 10^8 \text{ m} \qquad M = 6.0 \times 10^{24} \text{ kg}$$

and are asked to calculate the value v at the point $y = R$, with R the radius of the earth $\equiv 6.4 \times 10^3$ km. Using the known value for G, the quantity GM/a has the value

$$\frac{GM}{a} = 6.7 \times 10^{-11} \frac{\text{N-m}^2}{\text{kg}^2} \times \frac{6.0 \times 10^{24} \text{ kg}}{10^8 \text{ m}}$$
$$= 4.0 \times 10^6 \frac{\text{m}^2}{\text{s}^2}$$

and the parameter GM/R has the value

$$\frac{GM}{R} = 6.7 \times 10^{-11} \frac{\text{N-m}^2}{\text{kg}^2} \times \frac{6.0 \times 10^{24} \text{ kg}}{6.4 \times 10^6 \text{ m}} = 6.3 \times 10^7 \frac{\text{m}^2}{\text{s}^2}$$

Substitution into (5-11) yields (since $v_0 = 0$)

$$\frac{1}{2} v^2 - 6.3 \times 10^7 \frac{\text{m}^2}{\text{s}^2} = 0 - 4.0 \times 10^6 \frac{\text{m}^2}{\text{s}^2}$$

and this leads to

$$v = -1.1 \times 10^4 \text{ m/s}$$

5-7 Circular motion

According to the analysis of Section 3-10, if a particle of mass m travels with uniform speed v_0 around a circular path it will undergo a centripetal acceleration $\mathbf{a}_c$ which is directed radially inward from the instantaneous position of the particle. The magnitude of this acceleration is v_0^2/R with R the radius of the orbit. To sustain this uniform motion, it is necessary, according to the second law, to exert a force parallel to $\mathbf{a}_c$ and of magnitude mv_0^2/R.

The purpose of this section is to generalize these ideas to the case of nonuniform circular motion. For this case we find that in addition to its centripetal acceleration, the particle will also undergo a *tangential acceleration* $\mathbf{a}_t$, which is oriented along the direction of motion (see Figure 5-14). The net or total acceleration of a particle going along a circular orbit is therefore the vector sum $(\mathbf{a}_c + \mathbf{a}_t)$ of these two accelerations. Since $\mathbf{a}_c$ and $\mathbf{a}_t$ are mutually perpendicular, the magnitude of the total acceleration is $(a_c^2 + a_t^2)^{1/2}$. Just as a radial force is required to produce $\mathbf{a}_c$, a force tangent to the circle is required to produce $\mathbf{a}_t$.

Consider, in Figure 5-15, a particle traveling counterclockwise around a circle of radius R under the action of a force. Let us draw a reference line AB

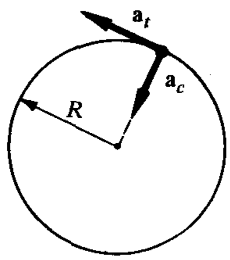


Figure 5-14

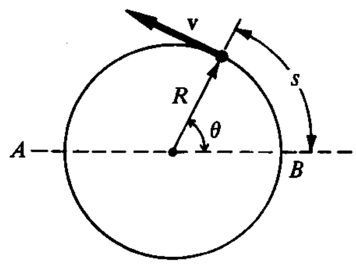


Figure 5-15

through the center of the circle and describe the instantaneous position of the particle at time t in terms of its angular displacement θ relative to this line. If B is selected to coincide with the initial position of the particle, then the angle θ represents its position at time t. Alternatively, we may describe the position of the particle at time t in terms of the circumferential distance s that it has traveled during this time interval. The relation between s and θ is simplified by expressing the angular displacement θ in units of angular measure of the radian (rad). For then, by definition of a radian, it follows that s and θ are related by

$$s = R\theta \tag{5-12}$$

As indicated in Figure 5-15, the velocity $\mathbf{v}$ of the particle as it goes around the circle is tangent to the circle at the instantaneous position of the particle. The argument used in Example 3-5 for the case of uniform speed applies equally well to the present case. To relate the magnitude of $\mathbf{v}$ to the circumferential displacement s, suppose that the particle undergoes a displacement Δs in a time interval Δt. Its speed v is, by definition, the limit approached by the ratio $\Delta s/\Delta t$ as $\Delta t \to 0$. It follows from the definition of the derivative that

$$v = \frac{ds}{dt} \tag{5-13}$$

so if the position $s \equiv s(t)$ of the particle is known at any time t, the speed v may be obtained by differentiation. Note the similarity between (5-13) and its one-dimensional analogue in (2-6).

If the speed v of the particle as it goes around the circle is *not* uniform, the particle is said to have a *tangential acceleration* $\mathbf{a}_t$. Applying the definition in (3-20) through (3-22) to this case, and using the arguments in Sections 3-9 and 3-10, we find that the total acceleration $\mathbf{a}$ of the particle consists of the sum of two terms: (a) a centripetal acceleration $\mathbf{a}_c$, which is directed radially inward and has magnitude v^2/R; and (b) a tangential acceleration $\mathbf{a}_t$, which is tangent to the circle and has magnitude

$$a_t = \frac{dv}{dt} \tag{5-14}$$

The directions of $\mathbf{a}_c$ and $\mathbf{a}_t$ are thus as shown in Figure 5-14. For the special

case of uniform motion, the derivative dv/dt vanishes and, consistent with our findings in Chapter 3, the particle has in this case only a centripetal acceleration. The similarity between (5-14) and its one-dimensional analogue should also be noted.

Besides $\mathbf{v}$, $\mathbf{a}_c$, and $\mathbf{a}_t$, there are two related quantities that also serve to characterize circular motion. These are: the *angular velocity* ω of the particle and its *angular acceleration* α. Mainly for future reference, let us define these quantities here.

To this end let us differentiate both sides of (5-12). Taking into account, (5-13) we may express the result

$$v = R\omega \tag{5-15}$$

where $\omega \equiv d\theta/dt$ is the angular velocity of the particle. Differentiating (5-15) in turn and making use of (5-14) we find in the same way that

$$a_t = R\alpha \tag{5-16}$$

where $\alpha \equiv d\omega/dt$ is the particle's *angular acceleration.* It was stated previously that if the position $s \equiv s(t)$ of a particle is known, v and a_t may be computed by differentiation in accordance with (5-13) and (5-14). In the same way, if the angular position $\theta \equiv \theta(t)$ is known, the angular velocity and acceleration may be computed from

$$\begin{aligned} \omega &= \frac{d\theta}{dt} \\ \alpha &= \frac{d\omega}{dt} \end{aligned} \tag{5-17}$$

The units of ω are radians per second, and those of α are radians per second squared. Note that the radian is the ratio of two lengths, and is therefore dimensionless.

Example 5-9 A particle travels along a circular path of radius $R = 30$ cm in such a way that its angular position θ relative to some reference line is

$$\theta = at^2 + bt$$

with $a = 0.5\ \text{rad/s}^2$ and $b = 3.0$ rad/s. Calculate:
(a) The angular and tangential velocity at time t.
(b) The angular and tangential acceleration at time t.
(c) The centripetal acceleration at $t = 2.0$ seconds.

Solution
(a) The substitution of the given form for $\theta(t)$ into the first equation of (5-17) yields

$$\omega = \frac{d\theta}{dt} = \frac{d}{dt}(at^2 + bt) = 2at + b$$

which, by the use of (5-15), leads to

$$v = R\omega = R(2at + b)$$

(b) Differentiating these forms for ω and v we obtain, by use of (5-14) and (5-17),

$$\alpha = 2a$$
$$a_t = 2aR$$

(c) The substitution of the form for v, derived in (a), leads to

$$a_c = \frac{v^2}{R} = \frac{[R(2at+b)]^2}{R} = R(2at+b)^2$$

which at $t = 2.0$ seconds has the numerical value $a_c = 7.5\ \text{m/s}^2$.

Example 5-10 If the particle in Example 5-9 has a mass m, what force is required to produce the given motion?

Solution Since the particle has both a tangential and a centripetal acceleration, the required force **F** will have a component F_c directed radially inward and a tangential component F_t. Using the results of Example 5-9, we find that

$$F_c = ma_c = mR(2at+b)^2$$
$$F_t = ma_t = 2maR$$

Thus F_t is constant, since a_t is, whereas F_c increases quadratically with t, since the speed of the particle increases linearly with time.

Example 5-11 Consider a bead of mass m that is free to move on a thin, circular wire of radius R. If initially the bead is given a push so that it starts to travel at a speed v_0, and if μ_k is the coefficient of sliding friction, calculate its speed at any subsequent time t. Neglect gravity.

Solution For the conditions stated, only the wire can exert a force on the bead. This force may be resolved into a normal component N directed radially inward, and a tangential component due to friction, oriented opposite to the direction of motion. Since the normal component N is responsible for the centripetal acceleration v^2/R and the frictional force produces the tangential acceleration dv/dt, it follows from Newton's second law that

$$N = \frac{mv^2}{R}$$

$$m\frac{dv}{dt} = -\mu_k N$$

where the minus sign is due to the fact that frictional forces are always directed opposite to the actual motion.

To obtain an explicit formula for v, let us eliminate N between these two relations. The result is

$$\frac{dv}{dt} = -\frac{\mu_k v^2}{R}$$

Dividing both sides by v^2, we note that the left-hand side is the derivative of $(-1/v)$,

and the right-hand side is the corresponding derivative of $(-\mu_k t/R)$. Hence (see Section 2-7)

$$-\frac{1}{v}=-\frac{\mu_k t}{R}+c$$

with c an integration constant. Since at $t=0$, $v=v_0$, it follows that $c=-1/v_0$, and thus

$$v=\frac{v_0}{1+(\mu_k v_0/R)t}$$

Physically, this tells us that the bead gradually slows down from its initial speed v_0, although it never comes completely to rest in a finite time. This is plausible since as v decreases, so does the normal force N. This implies that the force of friction, which is proportional to N, must also become steadily smaller.

5-8 Motion in a viscous medium

According to (5-8), near the earth's surface all bodies, regardless of their mass and size, travel with the same constant acceleration **g**. On the other hand, we know that if a feather and a small heavy metal sphere are dropped simultaneously, then they fall at radically different rates, thus proving that their accelerations are not the same. This seeming paradox can be resolved by noting that when an object travels through air, or any other material medium, it experiences a force tending to inhibit its motion. It is not hard to see the physical origin and the nature of this motion-retarding force. For when a body travels through the air, the air molecules strike it preferentially in a direction that opposes this motion, thus causing it to slow down. The resultant acceleration of the body is usually attributed to a so-called *viscous force.*

Experiment shows that for many physical situations the viscous force F acting on an object may be described phenomenologically by assuming that F varies as some power of its velocity. In the *Principia,* Newton considered the cases of F varying both linearly and quadratically with velocity. Experimentally, we find that for small objects moving through air, for $v \leqslant 25$ m/s, F varies approximately as the first power of v, while for values of v in the range $25 \text{ m/s} \leqslant v \leqslant v_s$, where v_s is the velocity of sound ($\cong 325$ m/s), F is proportional to v^2. Customarily, we speak of *Stokes' law of resistance* for the case in which $F \sim v$, whereas the case in which $F \sim v^2$ is known as *Newton's law of resistance.* In both cases, the direction of the viscous force is in a direction opposite to the velocity vector. Thus the Stokes form of the viscous force **F** is

$$\mathbf{F}=-b\mathbf{v} \tag{5-18}$$

with b a positive constant. Except for its explicit velocity dependence, a viscous force is very similar to the force of friction considered previously.

To illustrate some of the ideas involved, consider the specific case of a body of mass m moving in the uniform gravitational field of the earth and subject to

the viscous force in (5-18). Assuming, for the sake of simplicity, that the motion is confined to a vertical line, and that the positive sense of the y-axis is vertically downward (see Figure 5-16), the equation of motion is

$$m\frac{dv}{dt} = mg - bv \tag{5-19}$$

where $v \equiv dy/dt$ is the velocity, and the minus sign before the bv term reflects the fact that the viscous force is opposite in sense to the velocity. That is, when $v > 0$, the term $-bv$ is negative and thus represents an upward force, whereas if $v < 0$, $-bv > 0$ and the viscous force is directed downward.

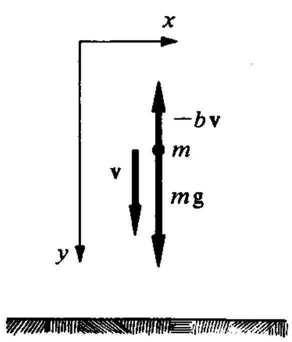


Figure 5-16

Imagine now an object, such as a raindrop, that is formed in the upper atmosphere and starts to fall downward from rest. At first it experiences only the gravitational force mg and thus accelerates downward under its action. As the raindrop now picks up speed, the viscous force comes into play, and the net or total force $(mg - bv)$ acting on the raindrop decreases steadily. Nevertheless, it continues to accelerate downward until finally its velocity is so great that the viscous force is equal and opposite to the gravitational attraction. At this point the raindrop ceases to accelerate, since there is no longer any net force acting on it. The velocity at which this occurs is called the *terminal velocity*, and we shall use the symbol v_T for it. On setting $v = v_T$ and the acceleration dv/dt in (5-19) to zero we find for the terminal velocity v_T

$$v_T = \frac{mg}{b} \tag{5-20}$$

The velocity with which a raindrop strikes the earth is normally its terminal velocity and were it not for the fact that the terminal velocity v_T for a raindrop is rather small, of the order of 2 m/s, life on earth would be very unpleasant indeed.

To obtain an explicit formula for the velocity of the body at any time t, let us multiply (5-19) by the factor $b\,dt/m(bv - mg)$. Making use of (5-20), we may express the result in the form

$$\frac{dv}{v - v_T} = -\frac{b}{m}\,dt$$

But the left-hand side is the differential $d[\ln(v - v_T)]$ and the right-hand side is the differential $d(-bt/m)$. Hence this formula may be expressed as

$$\ln(v - v_T) = -\frac{b}{m}t + c$$

with c an integration constant. Assuming that initially at $t = 0$, $v = v_0$, we find that c has the value $\ln(v_0 - v_T)$ and thus

$$\ln(v - v_T) = -\frac{b}{m}t + \ln(v_0 - v_T)$$

and this leads to the final form

$$v = v_T + (v_0 - v_T)e^{-bt/m} \tag{5-21}$$

Figure 5-17 shows a plot of v as a function of t for the case $v_0 < v_T$. Also shown is the corresponding zero-viscosity formula

$$v = v_0 + gt \qquad (b = 0) \tag{5-22}$$

Note that for times $t \ll m/b$, the particle moves as if there were no viscous force, whereas for times $t \gg m/b$, the particle travels essentially at its terminal velocity v_T. Thus the larger is the parameter b, the more rapidly does the falling object approach its terminal velocity.

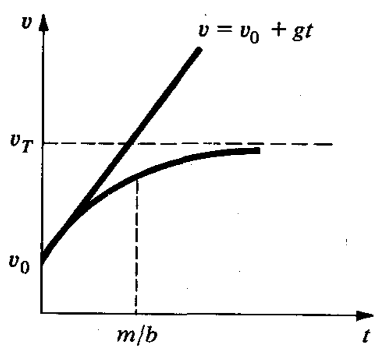


Figure 5-17

Example 5-12 A parachutist of mass 100 kg jumps from a plane when it is a distance of 2.5 km above the ground. As he nears the ground, he is observed to be descending at a constant speed of 12 m/s. Calculate the value of the parameter b and the time it takes him to achieve this terminal velocity.

Solution Since he is falling at constant speed, he has achieved his terminal speed. Substituting the values $v_T = 12$ m/s and $m = 100$ kg into (5-20) and solving for b, we find that

$$b = \frac{mg}{v_T} = \frac{100\text{ kg} \times 9.8\text{ m/s}^2}{12\text{ m/s}} = 82\text{ kg/s}$$

The time τ that it takes him to achieve this velocity can be estimated as the value at

which the exponent in (5-21) has the value unity—that is, when $b\tau/m = 1$ or, in other words, at time τ given by

$$\tau = \frac{m}{b} = \frac{100\text{ kg}}{82\text{ kg/s}} = 1.2\text{ s}$$

This means that after a few seconds the exponential term in (5-21) is negligible, and the parachutist falls to the ground at his terminal velocity of 12 m/s.

†5-9 The equivalence principle

In Chapter 4 we found that mass, as a physical entity, comes up in two distinct contexts. First, it occurs in the second law, $\mathbf{F} = m\mathbf{a}$, where it serves as a measure of the inertia—or the resistance to change in motion—of the body under consideration. Let us use the term *inertial mass* and the symbol m_I when we use mass in this sense. Second, mass occurs in Newton's law of universal gravitation

$$F = \frac{GMm}{r^2} \tag{5-23}$$

where in this case the mass m of the body is a measure of its ability to exert gravitational forces on other nearby bodies. When used in this second way, let us refer to the mass of the body as its *gravitational mass* and use the symbol m_G to represent it. In all discussions up to this point, it should be noted, we have implicitly assumed that m_I and m_G have precisely the same value. Let us consider briefly the evidence which justifies this assumption.

First, however, let us see how an experimental difference between the gravitational and the inertial mass of a body might be detected. To this end, consider a body moving in a uniform gravitational field $\mathbf{g}$. In terms of its gravitational mass m_G, the force on the body is $m_G\mathbf{g}$. Hence, according to the second law,

$$m_I\frac{d\mathbf{v}}{dt} = m_G\mathbf{g}$$

where m_I is the inertial mass of the body. On dividing both sides of this relation by m_I we find the equivalent form

$$\frac{d\mathbf{v}}{dt} = \left(\frac{m_G}{m_I}\right)\mathbf{g} \tag{5-24}$$

and thus if the ratio m_G/m_I were not the same for *all* bodies, then the accelerations of two bodies in the earth's uniform field would be different. This then is the basic idea underlying all experiments to measure the ratio m_G/m_I.

Sir Isaac Newton himself was the first to question this assumption of the

equality of m_I and m_G and to attempt to resolve it *experimentally*. In his own words,[2]

> But it has been long ago observed by others that (allowance being made for the small resistance of air) all bodies descended through equal spaces in equal times; and by the help of pendulums, that equality of times may be distinguished to great exactness.

He found that, to an accuracy of about one part in a thousand, m_G and m_I were equal. Subsequently, R. Eötvös, beginning in about 1890 and continuing on for about 25 years, verified this equality to an accuracy of five parts in 10^9. And even more recently, Roll, Krotkov, and Dicke have improved on this result and confirmed the equality of these two types of mass to one part in 10^{11}. Thus, as far as we can judge today, inertial and gravitational mass are one and the same; in particular, all bodies, regardless of their mass, undergo precisely the same acceleration in a uniform gravitational field.

Consider now, in Figure 5-18, an elevator whose supporting cable has snapped, and which is thus falling down the shaft at the acceleration **g**. Because of the equality of m_I and m_G, all objects inside the elevator will have precisely the same acceleration **g** as does the elevator itself. Hence if the observer O in the elevator releases an object, it will appear to him to go with uniform speed in a straight line. In other words, to the observer O it appears as if he is in a gravity-free inertial reference frame and the only forces acting are the contact forces associated with collisions with the walls. Of course, there is nothing unexpected or mysterious here, since to the stationary observer O' all bodies appear to be accelerating downward at the acceleration **g**, and it is only the relative accelerations between them that vanish. This effect is thus very similar to the feeling of weightlessness and related phenomena experienced by astronauts when in a stationary orbit about a planet.

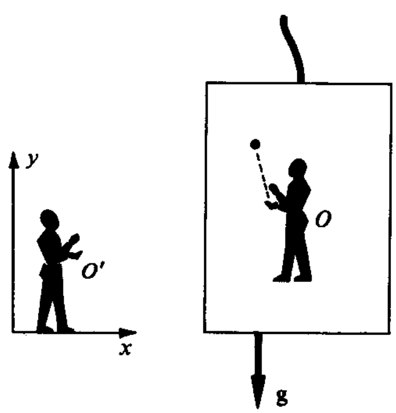


Figure 5-18

[2]As quoted in *Berkeley Physics Course*, vol. I, C. Kittel, W. D. Knight, and M. A. Ruderman. New York: McGraw-Hill, 1965, p. 437.

Taking note of this fact that the effects of a uniform gravitational field may be completely nullified by use of an appropriate accelerated reference frame, Einstein postulated his *equivalence principle*:

All freely falling (and nonrotating) laboratories are fully equivalent as far as the results of experiments carried out in them are concerned.

Or, equivalently, the laws of physics are the same in all such freely falling reference frames. Thus, an experiment carried out in a freely falling laboratory will yield the same results regardless of whether it is on earth or on Jupiter, or in a spaceship floating out in space. Note that if gravitational and inertial mass were not precisely equivalent, this principle of Einstein would be manifestly false.

One of the very interesting implications of the equivalence principle is that an ordinary light ray will be bent in a gravitational field. Consider, in Figure 5-19a, a flash of light emitted at a point A in a direction perpendicular to the wall of a freely falling elevator. Because of the equivalence principle, the resulting path of the light ray must be the same as if the elevator were nonaccelerating and very far away from all other matter. According to the laws of optics, the light ray will follow a straight line and reach the corresponding point B on the opposite wall. Indeed, this is precisely what the observer O fixed in the elevator sees. On the other hand, if this experiment is viewed from the viewpoint of a nonaccelerating observer O', we conclude that the light ray must bend. For, as shown in Figure 5-19b, since the system falls during the time interval that the light ray is in transit across the elevator and since the light ray is destined to strike the point B, it follows that the light ray must follow a curved path such as that shown in the figure. In effect then, *the light ray is bent in a gravitational field as if it were itself a material particle.* In other words, light rays also experience the gravitational field of large bodies.

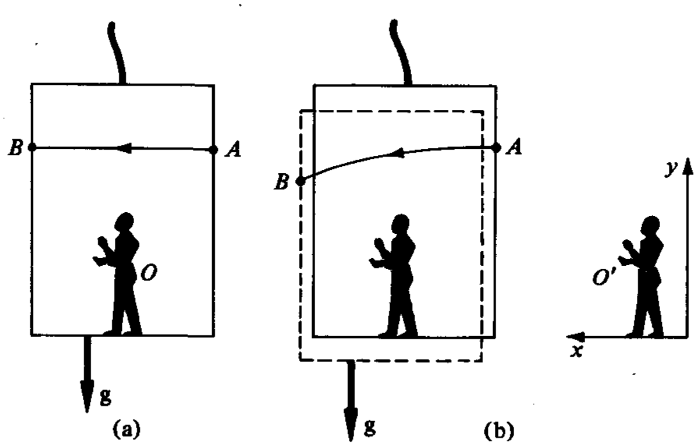


Figure 5-19

Because of the fact that the speed of light is so large, it is ordinarily very difficult to detect this gravitational bending of a light beam in the laboratory. However, it has been observed as a shift in the apparent position of a star when it appears in that part of the sky near the sun. Figure 5-20 shows a star at A sending out a light ray, which if it were not deviated by the sun's gravitational field would reach the observer's eye at B'. However, because of the solar gravitational field, the ray is bent and follows the path AB. Thus to the observer at B, the star appears to be located at the point A' which is displaced from its true position by a certain angle θ. This angle will be maximum when the light ray just grazes the sun, and calculations using the theory of relativity show that this maximum value (in radians) is

$$\theta = \frac{4GM_s}{R_s c^2} \tag{5-25}$$

where R_s and M_s are the radius and mass of the sun, respectively, and $c \cong 3 \times 10^8$ m/s is the speed of light. Substituting the known values for these parameters we find

$$\theta = 8.50 \times 10^{-6} \text{ rad} = 1.75'' = 0.00049^\circ$$

and this agrees, within 20 percent, with the value found by observation.

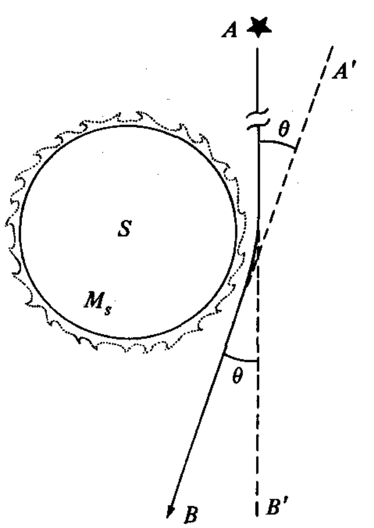


Figure 5-20

5-10 Summary of important formulas

In this chapter, Newton's second law for a particle

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} \tag{5-1}$$

has been studied for various physical systems, each with its own force function **F**. In certain cases it was possible to obtain explicit solutions for the trajectory of the particle $\mathbf{r}(t)$ at any time t.

QUESTIONS

1. Define or describe briefly the following terms: (**a**) static equilibrium; (**b**) coefficient of static friction; (**c**) uniform force field; (**d**) tangential acceleration; (**e**) angular velocity; and (**f**) viscous force.
2. Explain how you could use the principle of statics to determine your weight when you step on a (weight) scale and take a reading.
3. Suppose that a body is held in equilibrium by three forces $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$. Why must these three forces lie in a single plane? If it were held in equilibrium under the action of the four forces, $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$, $\mathbf{F}_4$, why need these not lie in a single plane?
4. Describe the nature of the force that causes you to accelerate when you suddenly start to run.
5. Give three illustrations of the role played by frictional forces in everyday life.
6. Describe in what sense the force of friction is a velocity-dependent force. In what sense is it velocity-independent?
7. Imagine yourself climbing up a rope. Describe in terms of the coefficient of friction why it is that the more tightly you hold the rope the less likely it is that you will slip.
8. Suppose you are given an inclined plane whose angle of elevation you can vary at will, and you wish to determine the relative value of μ_s between various objects and the plane. Describe an experiment to achieve this. (*Hint*: See Figure 5-8.)
9. Repeat Question 8, but this time devise an experiment to determine the coefficient of kinetic friction μ_k.
10. Are there any restrictions on the values that the coefficient of friction between two surfaces can assume? If so, what are they?
11. What is the physical basis of friction? Explain why there is no friction on a microscopic scale.
12. In selecting the origin to be at the center of the planet in Figure 5-12, why was it necessary to assume that the planet's motion could be neglected? What error would be made if the planet were in motion?
13. By what factor would your weight be less if you were a distance of one terrestrial radius above the earth's surface? Twice this distance?
14. An automobile is being driven around a circular track. What can you say about the direction of its acceleration if: (**a**) its speed is uniform; (**b**) it is accelerating; (**c**) it is decelerating?
15. Consider a bead that is forced to slide counterclockwise around a circular wire. If its angular acceleration is negative, state which of the following is negative and which is positive: (**a**) the tangential acceleration; (**b**) the angular velocity; (**c**) the centripetal acceleration; and (**d**) the linear velocity.
16. What is the physical mechanism underlying viscous forces? How would you expect the viscous force in air to vary with density? With temperature?
17. If a projectile goes through the air in

a direction parallel to that in which a wind is blowing, would you expect the force of the air on the object to be greater or less than it would be if there were no wind? Explain.

18. If a baseball travels at 5 m/s parallel to a 10-m/s wind, is there a viscous force on the baseball? What is the direction of the force that the air exerts on the ball?
19. Is the equation of motion in (5-19) for the body moving through a viscous medium invariant under Galilean transformations? Should it be? Explain.

PROBLEMS

1. A body is acted upon by the two forces $\mathbf{F}_1$ and $\mathbf{F}_2$ as shown in Figure 5-21. If $F_1 = 3.0$ newtons and $F_2 = 2.0$ newtons, find the magnitude and direction of the additional force that must be added so that the body will be in equilibrium.

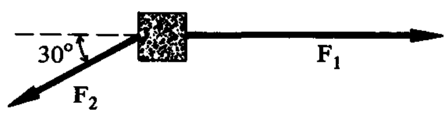


Figure 5-21

2. If a third force $\mathbf{F}_3$ of magnitude 2.0 newtons is applied to the body in Figure 5-21 and in a direction so that it makes an angle of 90° with $\mathbf{F}_1$ and an angle of 60° with $\mathbf{F}_2$, what additional force is required now to keep it in static equilibrium?
3. Three forces $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$ have magnitude 3.0 newtons, 4.0 newtons, and 5.0 newtons, respectively, and are applied to a body in such a way that it does not accelerate. What angles must the directions of these three forces make with one another?
4. A block of mass 2.0 kg is suspended from the ceiling of a room, as shown in Figure 5-22. Assume that the strings have negligible mass. **(a)** What must be the tension in the string labelled a? **(b)** Calculate the tensions in each of the other two strings.

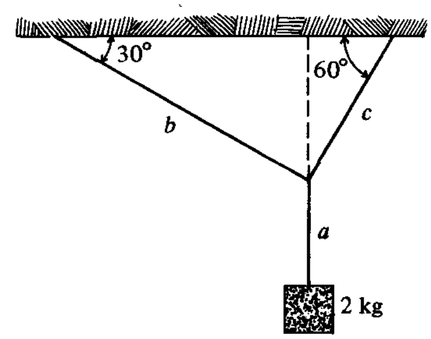


Figure 5-22

5. Calculate the tensions in each of the massless strings a, b, and c in Figure 5-23 in terms of the weight w of the block.

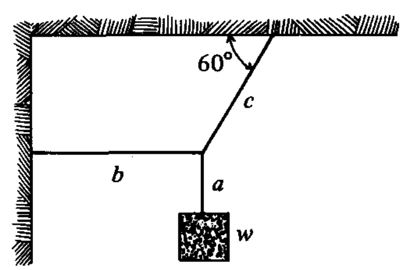


Figure 5-23

6. Figure 5-24 shows a block of mass 1.2 kg on an inclined plane of elevation angle 30° and connected to a second block of mass 0.2 kg by a massless string that passes over a smooth pulley.
   (a) Write the conditions for equilibrium of each block in terms of the tension T in the string and

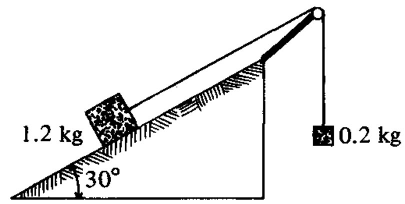


Figure 5-24

the force **R** that the plane exerts on the block.

(b) Calculate T.

(c) Calculate the components of **R** along and perpendicular to the plane.

7. A coal brick of mass 2.0 kg slides down a very dirty coal chute, with uniform speed. The angle of elevation of the chute is 60°. (a) What is the magnitude and direction of the force the chute exerts on the brick? (b) What is the coefficient of kinetic friction in this case?

8. Two small, electrically charged spheres, each of mass m, are suspended by two massless strings each of length l, as shown in Figure 5-25.

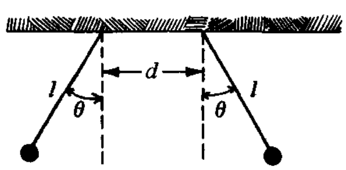


Figure 5-25

If θ is the angle each string makes with the vertical and if the force **F** between the particles acts along the line joining them, calculate (a) the strength of the force **F** and (b) the tension in each string.

9. Consider again the physical situation described in Figure 5-4. This time resolve the vectors along the axes of a coordinate system with the x-axis horizontal and the y-axis vertical.

(a) Show that the conditions for equilibrium in this system are

$$F \cos \alpha - N \sin \alpha - R \cos \alpha = 0$$

$$N \cos \alpha + F \sin \alpha - R \sin \alpha - mg = 0$$

(b) Show that N and R have precisely the same values as obtained by use of the coordinate system used in Example 5-3.

10. A block is given an impulsive push so that it travels across a rough horizontal surface at an initial speed of 5.0 m/s. (a) What is the coefficient of friction if the block comes to rest in a time interval of 5.0 seconds? (b) How far does the block travel before coming to rest?

11. A boy starts to slide across a sheet of ice. Assume that he travels a total distance of 30 meters in 25 seconds before coming to rest. (a) What was his initial speed? (b) What is the coefficient of friction?

12. A skier goes down a slope, which makes an angle α with respect to the horizontal. If μ_k is the coefficient of sliding friction between his skis and the slope, show that his acceleration a is

$$a = g(\sin \alpha - \mu_k \cos \alpha)$$

13. Making use of the result of Problem 12, if the slope is at an angle of 45° and is 100 meters in length, calculate his speed as he reaches the bottom (see Table 5-2 and assume that μ_k is independent of velocity): (a) If his hickory skis are waxed and the snow is dry. (b) If his hickory skis are waxed and the snow is wet.

14. A particle is projected down the inclined plane in Figure 5-26 with an initial speed of 2.0 m/s. Assume that the plane is smooth.

(a) What is its acceleration down the plane?

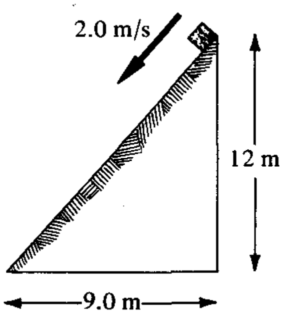


Figure 5-26

(b) How long after its release will it reach the bottom?
(c) What is its speed at the bottom?

15. Repeat Problem 14, but assume that this time there is a frictional force characterized by $\mu_k = 0.1$.
16. A particle is projected with an initial velocity v_0 down an inclined plane of angle α and of length l. What must be the coefficient of kinetic friction if the particle comes to rest just as it reaches the bottom of the plane?
17. For the physical situation described in Example 5-11 calculate, as explicit functions of time, the tangential and the centripetal accelerations of the bead and show that its total acceleration a is

$$a = \frac{v_0^2/R}{\left[1 + \frac{\mu_k v_0}{R}t\right]^2}[1 + \mu_k^2]^{1/2}$$

18. A bead of mass m is free to slide on a circular loop of wire of radius R. Assume that the plane of the ring is horizontal, that the coefficient of kinetic friction is μ_k, and that at $t = 0$ its velocity is v_0.
    (a) Show that the tangential acceleration is

$$\frac{dv}{dt} = -\mu_k\left[g^2 + \frac{v^4}{R^2}\right]^{1/2}$$

    with g the acceleration of gravity.
    (b) Find v as an explicit function of time, assuming that $v_0^2 \ll gR$.
19. To what elevation above the surface of the earth must you travel so that your weight drops to: (a) three fourths of its surface value; (b) one tenth of its surface value?
20. Using the data in Appendix B, calculate at what point along the line joining the earth and the moon you must be so that their gravitational forces on you will cancel.
21. For the case of a body originally at a distance $2R$ (where R is the radius of the earth) from the earth's center, make a plot of v against y, by use of (5-11), for the following values of v_0^2: $GM/3R$, $GM/2R$, and GM/R.
    (a) What is the maximum value that y can assume on each of these curves?
    (b) What is the physical distinction in your graphs between positive and negative values of v_0?
22. Show by use of (5-11) that if the initial speed v_0 is so large that $v_0^2 \geqslant 2GM/a$, then the body can escape from the influence of the planet. That is, show that it is possible for y to become arbitrarily large in these cases. The velocity $(2GM/a)^{1/2}$ is called the *escape velocity* from the planet.
23. By use of the result of Problem 22, calculate the escape velocity from the surfaces of (a) the earth; (b) our moon; and (c) the sun.
24. A spaceship is orbiting the earth at an elevation of 10^3 km above the surface of the earth when it explodes. Suppose that one of the larger fragments of the debris has an initial speed of 5×10^3 m/s directed radially *away* from the center of the earth.
    (a) What is the maximum distance from the center of the earth that this fragment reaches?
    (b) With what speed does it eventually strike the earth? (*Note*: In addition to this velocity in the

vertical direction, the fragment will also appear to have motion in the horizontal direction (to a terrestrial observer) because of the rotation of the earth about its own axis.)

25. Consider a point on a wheel that rotates at 10 revolutions per second. If the radius of the wheel is 0.3 m, calculate (a) its angular velocity; (b) its tangential velocity; and (c) its centripetal acceleration.

26. Consider a man of mass 80 kg who lives at the equator.
    (a) What is his angular velocity in radians per second due to the earth's rotation?
    (b) What is his centripetal acceleration?
    (c) What force is required to produce the acceleration in (b)? What is the origin of this force?

27. A particle of mass m is forced to go around a circle of radius R so that its position on the rim at time t is

$$s = \tfrac{1}{2}a_0t^2 - v_0t$$

with a_0 and v_0 positive constants. Calculate (a) its tangential acceleration at time t; (b) its angular acceleration at time t; and (c) the magnitude of the total force required at time t to sustain its motion.

28. Figure 5-27 shows an automobile of mass m traveling at a uniform speed v_0 on a curved road of radius R and banked at an angle α. The center of curvature is to the left and the vehicle is assumed to be traveling perpendicular and into the plane of the paper.
    (a) Show that the components of $\mathbf{F} = m\mathbf{a}$ along the horizontal and the vertical are given, respectively, by

$$V \sin\alpha - H\cos\alpha = \frac{mv_0^2}{R}$$

$$V\cos\alpha + H\sin\alpha = mg$$

    where V and H are the components of the force that the road exerts on the automobile.
    (b) Calculate the values for V and H and show that if

$$\tan\alpha = \frac{v_0^2}{Rg}$$

    then the component H of this force vanishes. Can you think of any reason why it might be desirable for the force to be completely normal to the road?

29. Making use of the result of (b) in Problem 28, calculate the angle α at which a road should be banked if it has a radius of curvature of 400 m and it is designed for automobiles traveling at 90 km/hr.

30. A commerical jet is traveling at a speed of 900 km/hr when the pilot banks the plane at a certain angle α as he makes a left turn in a circle of radius $R = 10$ km (see Figure 5-28). If the force that the floor exerts on a standing passenger is perpendicular to the floor, find the value of α. (*Hint*: Use the method of Problem 28.)

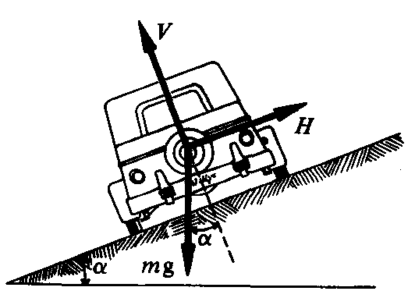


Figure 5-27

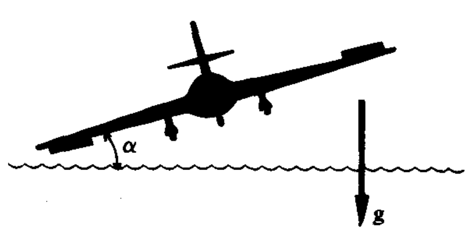


Figure 5-28

31. An object of mass 200 kg is dropped from a plane which is cruising at an elevation of 500 meters. Assuming that the viscous-force parameter b has the value $b = 250$ kg/s, calculate (a) the terminal velocity v_T and (b) the time after being released when it has achieved a velocity of $\frac{1}{2}v_T$. (c) Make a plot of v as a function of time.
32. Repeat Problem 31, but assume this time that at the instant when the object is dropped from the plane it is moving vertically upward at a speed of 50 m/s.
33. If the viscous force F acting on a falling body has the quadratic form

$$F = -\alpha v^2$$

with α a positive constant, show that the terminal velocity v_T in a uniform field g is

$$v_T = \sqrt{\frac{mg}{\alpha}}$$

*34. If the viscous force F has the form

$$F = -bv - \alpha v^2$$

calculate the associated value for the terminal velocity of a body of mass m falling in a uniform gravitational field in terms of the positive constants α, b, and m.

35. Suppose that a raindrop of mass 0.02 kg is formed at an elevation of 1 km above the ground.
    (a) If viscous effects were negligible, what would be its velocity on striking the ground?
    (b) In actual fact, it is found to strike the ground with its terminal velocity of approximately 4.0 m/s. What is the value of the parameter b associated with its motion? Assume the validity of (5-18).
36. In Example 5-12, suppose that at the instant that the parachutist leaves the plane it is in a downward dive of 15 m/s. Determine his velocity at any subsequent time t.

*37. (a) Show by integrating (5-19) directly that, at any time t,

$$mv = mgt - by + bh + mv_0$$

where h and v_0 are the initial values for y and v, respectively.

(b) Show by use of (5-21) that at any time t

$$y(t) = h + tv_T + \frac{m}{b}(v_0 - v_T) \times (1 - e^{-bt/m})$$

(c) By differentiation, show that the result of (b) satisfies (5-19).

38. Apply the result of (b) in Problem 37 to the case of the falling raindrop in Problem 35.
    (a) Make a plot of $y(t)$.
    (b) Through what distance has the raindrop fallen when its velocity is $\frac{1}{3}v_T$? When it is $\frac{1}{2}v_T$?

6 The harmonic oscillator

The thing that hath been is that which shall be; and that which is done is that which shall be done; and there is no new thing under the sun.

ECCLESIASTES

6-1 Periodic motion

Under certain conditions a physical system will undergo a type of motion that repeats itself at regular intervals. We call such motion *cyclic* or *periodic* and the time it takes the system to execute a complete cycle is called the *period.* The motion of the center of the earth about the sun, for example, is periodic with a period of one year. Similarly, the motion of a fixed point on the earth's equator relative to its center is periodic with a period of one day, and the second hand of a watch undergoes periodic motion with a period of one minute. If a body, such as the bob of a pendulum, moves periodically back and forth over the same path, then we also describe the motion by calling it *oscillatory.* Figure 6-1 shows several physical systems which undergo periodic motion.

One of the very important parameters used to characterize any type of periodic motion is the period T. If the motion of a system is periodic and is repeated in a time interval of length τ, then it will also be repeated in time intervals of length $2\tau, 3\tau, \ldots$. Hence, to give a unique meaning to the period T of such motion, we shall define it to be the *shortest* time interval for the

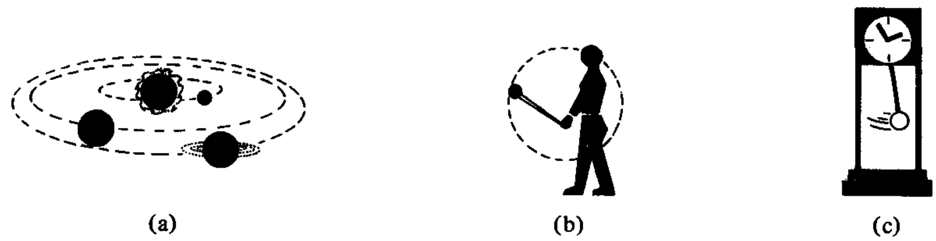


Figure 6-1

motion to be repeated. This implies, for example, that if $\mathbf{r}(t)$ and $\mathbf{v}(t)$ are, at time t, the position and velocity of a particle undergoing periodic motion, then the period T is the *shortest* time interval that must elapse before the position and velocity again assume the values $\mathbf{r}(t)$ and $\mathbf{v}(t)$ respectively. Note that in order for T to be the period, both the position and velocity of the particle must have the same values at time $(t + T)$ as they do at time t. If one of these conditions is not satisfied, then T is *not* the period.

A second important parameter used to characterize periodic motion is the *frequency*. We define the frequency ν by the formula

$$\nu = \frac{1}{T} \tag{6-1}$$

Thus ν represents the number of cycles per second or the number of times per second that the motion is repeated. The conventional unit of frequency is the hertz (Hz), which is defined as one complete repetition of motion per second. Thus, if the period of a certain motion is 0.25 second, then, by (6-1),

$$\nu = \frac{1}{T} = \frac{1}{0.25\text{ s}} = 4.0\text{ Hz}$$

and this means that the motion repeats itself 4 times each second.

6-2 Harmonic motion

Consider a body which is acted upon by certain forces so that it oscillates periodically back and forth through an equilibrium point. We say that the body is undergoing *harmonic motion*, provided that the variation in time of the position and velocity of the body are described by the trigonometric functions of the sine or the cosine. As we shall see, the motion of a particle attached to a vibrating spring, the small-amplitude oscillations of a pendulum, as well as the surge of electric charge in an LC circuit are all examples of physical situations involving harmonic motion.

Figure 6-2 shows a plot of $\sin\theta$ and $\cos\theta$ as functions of the angle θ in radians. Note that each of these trigonometric functions is periodic, with

period 2π, and thus

$$\begin{aligned} \sin(\theta + 2\pi) &= \sin\theta \\ \cos(\theta + 2\pi) &= \cos\theta \end{aligned} \tag{6-2}$$

for any angle θ. Let us now see how periodic motion can be described by the use of these functions.

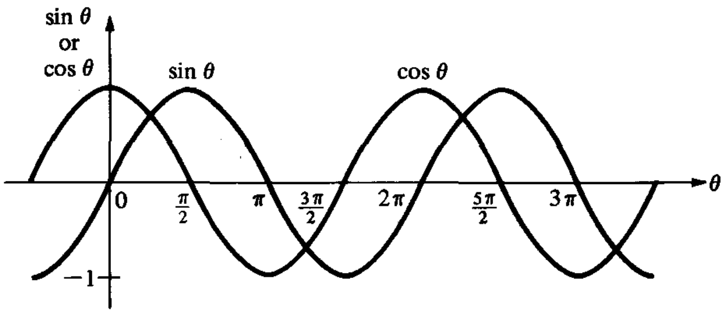


Figure 6-2

To this end, consider the function of time, $f(t)$, defined by

$$f(t) = \sin(\omega t + \alpha) \tag{6-3}$$

where α is a dimensionless constant and ω is a positive parameter with units of seconds^{-1}. At a later instant $(t + T)$, $f(t)$ assumes the form

$$f(t + T) = \sin[\omega(t + T) + \alpha] = \sin[(\omega t + \alpha) + \omega T]$$

Therefore, in order for $f(t)$ to be periodic with period T, that is, for $f(t + T) = f(t)$, it follows from (6-2) that it is necessary that

$$\omega T = 2\pi \tag{6-4}$$

In other words, the function $f(t)$ in (6-3) is periodic with period $T = 2\pi/\omega$. Some alternate but equivalent forms for $f(t)$ are

$$f(t) = \sin\left(\frac{2\pi}{T}t + \alpha\right) = \sin(2\pi\nu t + \alpha) \tag{6-5}$$

Figure 6-3 shows a plot of $f(t)$ as a function of t for the special choice $\alpha = 0$. Note that $f(t)$ has maxima at

$$t = \frac{T}{4}, \frac{5T}{4}, \ldots \left(= \frac{\pi}{2\omega}, \frac{5\pi}{2\omega}, \ldots\right)$$

and minima at

$$t = \frac{3T}{4}, \frac{7T}{4}, \ldots$$

and vanishes at half-integral multiples of T.

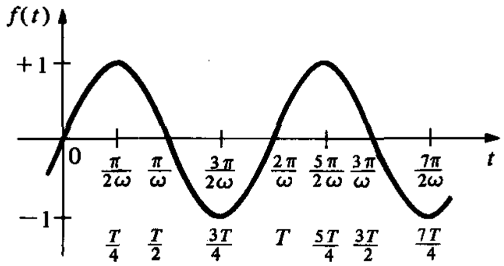


Figure 6-3

If in (6-3) the replacement $\alpha \to \alpha + \pi/2$ is made, then since

$$\cos \theta = \sin \left(\theta + \frac{\pi}{2}\right)$$

it follows that the function $g(t)$ defined by

$$g(t) = \cos(\omega t + \alpha) \tag{6-6}$$

is also periodic with the same period T in (6-4).

Example 6-1 A particle travels along the y-axis of a certain coordinate system so that at time t its position is

$$y(t) = y_0 \sin\left(\frac{2\pi}{3} t + \alpha\right)$$

where y_0 and α are fixed constants and t is measured in seconds. What is the period and the frequency of the motion?

Solution Comparing $y(t)$ with (6-5), we see that the period T is

$$T = 3 \text{ s}$$

and therefore, according to (6-1), the frequency ν is

$$\nu = \frac{1}{T} = \frac{1}{3\text{ s}} = 0.33 \text{ Hz}$$

Example 6-2 Find the period and the frequency of a particle whose position is described by

$$y(t) = A \cos\left(17t + \frac{\pi}{3}\right)$$

with A a fixed constant and t measured in seconds.

Solution Comparing this formula for $y(t)$ with (6-6) and utilizing (6-4), we find that

$$\frac{2\pi}{T} = 17 \text{ s}^{-1}$$

that is,

$$T = \frac{2\pi}{17\text{ s}^{-1}} = 0.37 \text{ s}$$

Thus ν has the value

$$\nu = \frac{1}{T} = \frac{1}{0.37\text{ s}} = 2.7 \text{ Hz}$$

6-3 Hooke's law

Of the many physical systems that are known to undergo harmonic motion, for purposes of detailed study let us select one of the simplest of these, the *simple harmonic oscillator.* It consists of a small body attached to one end of a spring, the other end of which is held firmly in place. On activating the system by displacing the particle from its equilibrium position and then releasing it, we find as a consequence of the resultant compressing and stretching of the spring that the particle executes an oscillatory, periodic motion of a certain type. It will be shown below that this motion is not only periodic, but also harmonic; and it is for this reason that we refer to it as *simple harmonic motion* (SHM). Before going into details, however, let us examine first the characteristics of the force that produces SHM and some of its qualitative aspects.

Consider, in Figure 6-4a, a massless spring of natural length l, hanging vertically from a ceiling. In order to measure the force that the spring can exert when it is stretched, let us carry out a sequence of experiments by successively attaching various bodies to its lower end and in each case lowering the body gradually until it achieves its equilibrium position (Figure 6-4b). These measurements show that the distance d by which a body of mass m stretches the spring from its equilibrium length l is directly proportional to the weight mg of that body. Calling the coefficient of proportionality $1/k$, we may write

$$mg = kd \tag{6-7}$$

where k is a positive constant characteristic of the spring, called the *spring constant.* If, for example, a 1.5-kg mass stretches a spring by 1 cm, then a 3-kg body will stretch it by 2 cm, and the spring constant for this spring is

$$k = \frac{mg}{d} = \frac{1.5\,\text{kg} \times 9.8\,\text{m/s}^2}{10^{-2}\,\text{m}} = 1.5 \times 10^3\,\text{N/m}$$

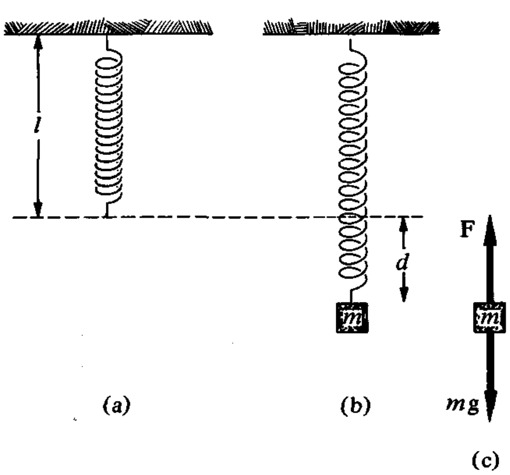


Figure 6-4

In general, the larger the value of k the stiffer is the spring, and the larger must be the weight of the attached body to produce a given elongation.

In order to express the experimental result in (6-7) in a more convenient way, consider again the situation in Figure 6-4. In part (c) of the figure, the body has been isolated and the two forces acting on it are shown by arrows. These forces are: the force of gravity $m\mathbf{g}$, which acts vertically downward, and the force $\mathbf{F}$ that the spring exerts on the body. The requirement that the body be in static equilibrium is that $\mathbf{F}$ be equal and opposite to the weight $m\mathbf{g}$ of the body. Thus the direction of $\mathbf{F}$ is vertically upward and, according to (6-7), its magnitude is

$$F = kd \tag{6-8}$$

In other words, the force $\mathbf{F}$ on the particle is proportional to the distance by which the spring is stretched and is directed in such a way that the spring tends to contract to its unstretched equilibrium length. Similarly, if the spring is compressed, the force $\mathbf{F}$ it exerts has the magnitude in (6-8) and is directed in such a way that the spring tends to expand to its equilibrium state. In either case, the spring has a tendency to regain its natural unstretched length l.

The results of these experiments may be summarized in the following way. If an agent stretches a spring from its equilibrium length by a displacement $\mathbf{r}$, then the force $\mathbf{F}$ that the spring exerts on the agent is

$$\mathbf{F} = -k\mathbf{r} \tag{6-9}$$

where k is the spring constant of the spring. Since k is an inherently positive constant, it follows that the direction of $\mathbf{F}$ is always such as to cause the spring to return to its equilibrium length. If the spring is compressed, the direction of $\mathbf{F}$ is such as to make it expand, and if it is stretched, $\mathbf{F}$ is directed so that the spring tends to contract. The relation in (6-9) is known as *Hooke's law*, and is found to be valid not only for springs but also for a wide variety of elastic materials, provided that the material is not stretched beyond what is known as its *elastic limit.*

For the special case of one-dimensional motion, Hooke's law assumes the component form

$$F = -kx \tag{6-10}$$

where the body has been assumed to be confined to motion along the x-axis and the origin has been selected to coincide with the equilibrium position of the moveable end of the spring.

Example 6-3 Figure 6-5a shows a block attached to two springs of constants k_1 and k_2. Assuming that initially both springs have their natural lengths l_1 and l_2, respectively, so that neither exerts a force on the block, what is the force on the block if it is displaced a distance d to the right? Assume that the surface on which the block rests is smooth and horizontal.

Solution Figure 6-5b shows the situation when the block has been displaced a distance d to the right. The spring on the left is stretched a distance d, and thus it exerts on the block a force k_1d directed to the left. Similarly, the other spring is compressed

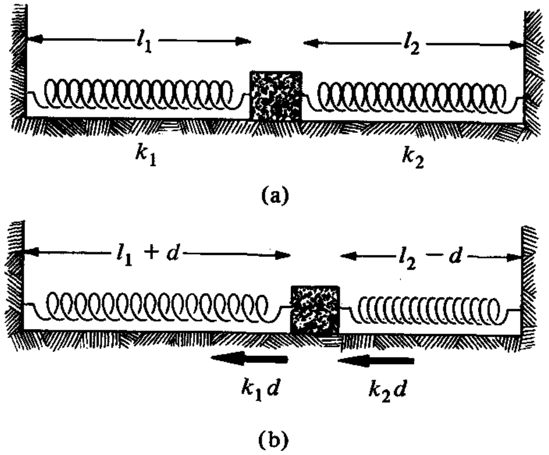


Figure 6-5

by the same distance d, and thus it exerts on the block a force of strength k_2d also directed to the left! Thus, the total force **F** on the block which is the vector sum of these is directed to the left, and has the magnitude

$$F = (k_1 + k_2)d$$

Just as for the case of a single spring then, when the block is displaced, it tends to return to its equilibrium position, in which both springs are in their natural relaxed state.

6-4 Equation of motion

Consider, in Figure 6-6, a block of mass m confined to motion along a smooth, horizontal surface and attached to one end of a spring of constant k and of natural length l. Assume that the spring is horizontal and that its other end is fixed to a vertical wall. To describe the motion of this oscillator, let us set up a coordinate system with origin at the equilibrium position at a perpendicular distance l from the wall and with the x-axis horizontal. The only force on the block that is capable of influencing its horizontal motion is the force due to the spring. Hence, it follows from Hooke's law that when the block is at a distance x from its equilibrium position, the restoring force **F** acting on it has only the nonzero component $F = -kx$. Substituting this into Newton's second law, we obtain

$$m\frac{dv}{dt} = -kx \tag{6-11}$$

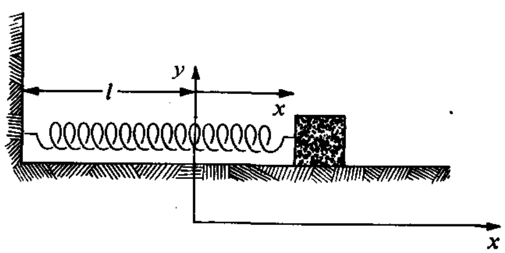


Figure 6-6

where $v \equiv dx/dt$ is the velocity of the block. Alternatively, this may be expressed in the form

$$m\frac{d^2x}{dt^2} = -kx \tag{6-12}$$

where d^2x/dt^2 is the second derivative of x, that is, the derivative of dx/dt. Note that (6-11) predicts that the acceleration dv/dt of the block in Figure 6-6 is directed to the left if the block is to the right of the equilibrium position and, conversely, it is directed to the right when the block is displaced to the left. In other words, (6-11) is consistent with the fact that regardless of the location of the block, the force of the spring tends to return the block toward the equilibrium position at $x = 0$.

Although the equation of motion of the simple harmonic oscillator in (6-11) has been derived only for the special case of horizontal motion, it is easy to see that (6-11) is much more generally valid. Consider again the physical situation in Figure 6-4b, in which a block of mass m is in equilibrium at a distance $(l + d)$ below the point of suspension of a spring of natural length l. Let us set up a coordinate system as in Figure 6-7, with the origin at the equilibrium position and with the y-axis vertical. At an instant when the block is a distance y above this equilibrium point, the force $\mathbf{F}$ on the particle is

$$\mathbf{F} = -\mathbf{j}mg + \mathbf{j}k(d - y)$$

where the first term represents the downward force of gravity and the second represents the spring force, with magnitude proportional to the elongation of the spring $(d - y)$ from its equilibrium length l. However, because of (6-7), the terms $-mg$ and kd cancel. Hence the total force acting on the block in Figure 6-7 is simply $-\mathbf{j}ky$, and the equation of motion is

$$m\frac{dv}{dt} + ky = 0$$

with $v \equiv dy/dt$ in this case. Comparison with (6-11) shows that the equation of motion in (6-11) is again applicable provided that the symbol x there refers to the displacement of the block from equilibrium.

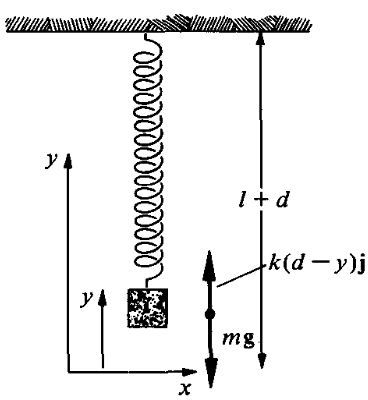


Figure 6-7

Example 6-4 A block of mass m is attached to a spring of constant k and natural length l and rests on a smooth, inclined plane of angle α (Figure 6-8).

(a) At equilibrium, by what amount d is the spring stretched?

(b) In terms of an x-axis, with origin at the equilibrium position and pointing down the plane, what is the equation of motion for the block if its equilibrium position is disturbed?

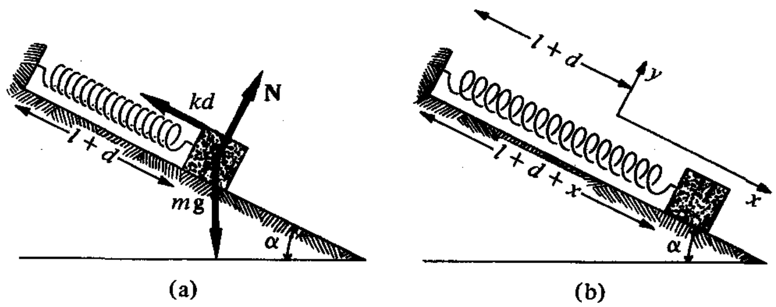


Figure 6-8

Solution

(a) Figure 6-8a shows the situation at equilibrium. Since there is no motion along the plane, it follows that the upward pull kd of the spring on the block is equal and opposite to the component $mg \sin \alpha$ of its weight mg along the plane. Hence

$$mg \sin \alpha = kd$$

and this determines the distance of stretching, d, of the spring.

(b) Figure 6-8b shows the situation if the block is displaced an additional distance x down the plane. Since the spring is now stretched by the amount $(x + d)$, it follows that it exerts on the block a force $k(x + d)$ directed upward along the plane. As before, the force of gravity has a component $mg \sin \alpha$ that acts downward along the plane. Thus the total force F_x acting along the plane is

$$\begin{aligned} F_x &= mg \sin \alpha - k(x + d) \\ &= -kx \end{aligned}$$

where the second equality follows by use of the result of (a). Substituting this formula for F_x into the second law we find that (6-11) is again applicable.

Since there is no motion normal to the plane, it follows that $N = mg \cos \alpha$.

6-5 Qualitative aspects of the motion

Before proceeding in the next section with the solution of (6-11) to obtain x as an explicit function of time, let us consider some qualitative aspects of the motion implied by this relation. As for the corresponding case of a particle moving in an inverse-square force field, we shall do this by integrating (6-11) to obtain a relation between v and x.

To this end, let us multiply both sides of (6-11) by $v = dx/dt$. The result is

$$mv\frac{dv}{dt} = -kx\frac{dx}{dt} \tag{6-13}$$

Now since the left-hand side is the derivative of $mv^2/2$ and since the derivative of x^2 is $2x\,dx/dt$, it follows that (6-13) may be expressed equivalently as

$$\frac{1}{2}mv^2 = -\frac{1}{2}kx^2 + c \tag{6-14}$$

with c a constant of integration. If we define the *amplitude* A of the motion to be the value for $|x|$ at the instant when $v = 0$, then c will have the value $\frac{1}{2}kA^2$. Thus (6-14) may be expressed as

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kA^2 \tag{6-15}$$

and this is the analogue of (5-11) for motion in the inverse-square gravitational field. Just as in that case, many interesting questions of physical interest can be answered directly by its usage. For reasons to be discussed in Chapter 8, (6-15) is known as the *energy integral* of the harmonic oscillator.

Consider, for example, the situation in Figure 6-9, where a block attached to a spring is free to move on a smooth, horizontal surface. Suppose that initially the block is displaced to the right by a certain distance A and then released. As the spring starts to contract, the block will move to the left and its position x and velocity v at any time t will vary but in such a way that (6-15) is satisfied at each instant. Thus as the block travels to the left, x decreases, and it follows that the velocity v must increase steadily until the block is at the equilibrium position at $x = 0$. As shown in Figure 6-9b, here, according to (6-15), the velocity v is a maximum and has the value

$$v_m = A\sqrt{\frac{k}{m}} \tag{6-16}$$

directed to the left. Even though there is no force on the block when it is at $x = 0$, since it has at this point the velocity v_m to the left, it will persist in its motion and continue its journey right through the equilibrium point. As it now travels to the left of the equilibrium point, its displacement x increases in magnitude and, in accordance with (6-15), its velocity now decreases. Eventually the block arrives at a point where it is again at rest. Substituting the

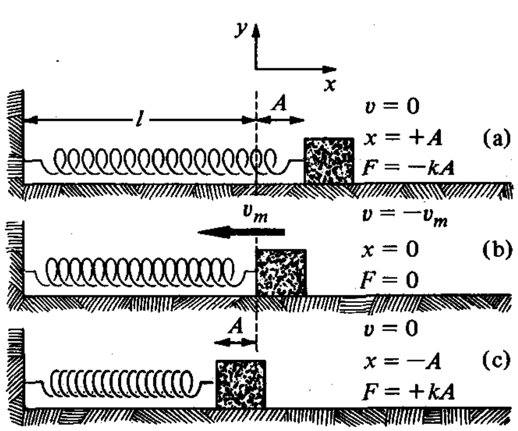


Figure 6-9

value $v = 0$ into (6-15), we find that this point corresponds to the value $x = -A$. In other words, as shown in Figure 6-9c, the block, which started out at rest at the point $x = A$, also comes to rest at the point $x = -A$ at a distance A on the other side of the equilibrium position. During the time that the block goes from $x = A$ to $x = -A$, the spring is compressed from its original length $(l + A)$ to its final length $(l - A)$. As it subsequently uncoils, it forces the block to move to the right in a manner precisely analogous to its original motion to the left from the point $x = +A$.

It follows from this discussion that the block oscillates back and forth between the points $x = +A$ and $x = -A$ and that its motion is periodic. At these points $x = \pm A$, the velocity vanishes while the magnitude of the Hooke's law force is a maximum with the value kA. Correspondingly, the magnitude of the acceleration of the block achieves its maximum value kA/m at these points. At the equilibrium position, $x = 0$, the force and the acceleration vanish, while the magnitude of the velocity achieves its maximum value $A\sqrt{k/m}$.

Example 6-5 Suppose that the block in Figure 6-9 has a mass of 50 grams and is initially displaced a distance 2.0 cm and then released. If the spring constant of the spring is 0.8 N/m, what is the velocity of the particle when:

(a) It goes through the equilibrium point?

(b) It is 1.0 cm from the equilibrium point?

Solution The parameter values are, in this case, $A = 2.0 \times 10^{-2}$ meter, $m = 0.05$ kg, and $k = 0.8$ N/m.

(a) When the particle goes through the equilibrium point, its velocity according to (6-16) has the maximum value

$$v_m = A\sqrt{\frac{k}{m}} = 2.0 \times 10^{-2}\text{ m}\sqrt{\frac{0.8\text{ N/m}}{0.05\text{ kg}}} = 0.08\text{ m/s}$$

(b) Solving (6-15) for v, we obtain

$$v = \sqrt{\frac{k}{m}}(A^2 - x^2)^{1/2}$$

and substituting the known values for k, m, and A and the value $x = 10^{-2}$ meter, we find when the block is 1 cm from the equilibrium position that its velocity is

$$v = \sqrt{\frac{k}{m}}(A^2 - x^2)^{1/2} = \sqrt{\frac{0.8\text{ N/m}}{0.05\text{ kg}}}(4.0 \times 10^{-4}\text{ m}^2 - 10^{-4}\text{ m}^2)^{1/2}$$
$$= 0.069\text{ m/s}$$

6-6 Solution of the equation of motion

One way to solve the equation of motion for the simple harmonic oscillator in (6-11) is to integrate the energy integral in (6-15) directly. However, rather than follow that route it is instructive to proceed differently by use of the rules for differentiating trigonometric functions. The formulas of interest in this

connection are derived in Appendix C and are:

$$\frac{d}{dt}\sin(\omega t+\alpha)=\omega\cos(\omega t+\alpha) \tag{6-17}$$

$$\frac{d}{dt}\cos(\omega t+\alpha)=-\omega\sin(\omega t+\alpha) \tag{6-18}$$

where α and ω are arbitrary, but fixed, time-independent constants.

Since in the equation of motion in (6-11) the parameters k and m appear only in the ratio k/m, it is convenient to define a parameter ω by

$$\omega=\sqrt{\frac{k}{m}} \tag{6-19}$$

The equation of motion, say in the form in (6-12), may then be expressed as

$$\frac{d^2x}{dt^2}=-\omega^2x \tag{6-20}$$

and the main result of this section is that the solution of this relation is

$$x(t)=A\cos(\omega t+\alpha) \tag{6-21}$$

with A and α arbitrary constants. The physical significance of A and α, which are known respectively as the *amplitude* and the *phase* of the motion, as well as that of the parameter ω, will be discussed in Section 6-7.

To confirm that (6-21) satisfies (6-20), let us differentiate the given form for $x(t)$. The result is

$$v=\frac{dx}{dt}=\frac{d}{dt}[A\cos(\omega t+\alpha)]=-\omega A\sin(\omega t+\alpha) \tag{6-22}$$

where use has been made of (6-18) and the fact that A, α, and ω are constants. A second differentiation then yields the acceleration:

$$\begin{aligned}\frac{d^2x}{dt^2}=\frac{dv}{dt}&=\frac{d}{dt}[-\omega A\sin(\omega t+\alpha)]\\&=-\omega^2A\cos(\omega t+\alpha)=-\omega^2x\end{aligned}$$

where the last equality follows by use of (6-21). Substituting this formula for the acceleration and the stated form for $x(t)$ into (6-20), we see that (6-21) indeed satisfies the equation of motion in (6-20) for *any* choice of the constants A and α.

Further confirmation of this fact can be obtained by substituting (6-21) into the energy integral in (6-15). Making use of (6-22), we find that

$$\begin{aligned}\frac{m}{2}v^2+\frac{1}{2}kx^2&=\frac{m}{2}\omega^2A^2\sin^2(\omega t+\alpha)+\frac{1}{2}kA^2\cos^2(\omega t+\alpha)\\&=\frac{1}{2}kA^2[\sin^2(\omega t+\alpha)+\cos^2(\omega t+\alpha)]\\&=\frac{1}{2}kA^2\end{aligned}$$

where the second equality follows since $m\omega^2 = k$ according to (6-19), and the last since $\sin^2\theta + \cos^2\theta = 1$ for all θ. Hence, not only does the solution in (6-21) satisfy the energy integral, but in addition we find that the arbitrary constant A in (6-21) has the physical significance of being the *amplitude* of the motion.

Example 6-6 Show that another form for the solution of (6-20) is

$$x(t) = x_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t \tag{6-23}$$

and give a physical interpretation for the constant parameters x_0 and v_0.

Solution Differentiating (6-23) we find, by use of (6-17) and (6-18), that

$$v = \frac{dx}{dt} = -\omega x_0 \sin \omega t + v_0 \cos \omega t$$

and

$$\begin{aligned} \frac{d^2x}{dt^2} &= -\omega^2 x_0 \cos \omega t - v_0\omega \sin \omega t \\ &= -\omega^2 \left[x_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t \right] \end{aligned}$$

Substituting back into (6-20) we find that (6-23) is indeed a solution of the equation of motion for any choice of the constants x_0 and v_0.

Since $\cos 0° = 1$ and $\sin 0° = 0$, it follows by setting $t = 0$ in (6-23) that $x(0) = x_0$. Hence x_0 is the position of the particle at $t = 0$. Similarly, by setting $t = 0$ in the above formula for the velocity $v(t)$, we find that v_0 is the initial velocity of the particle.

6-7 Physical significance of the parameters

The solution (6-21) of the equation of motion for the simple harmonic oscillator involves the three parameters A, α, and ω. Of these, only ω, which is defined in (6-19), is involved directly in the equation of motion; the amplitude A and the phase α have to do *not* with an intrinsic property of the oscillator but only with the initial conditions.

Let us first consider the parameter ω. From the discussion of Section 6-5 we know that SHM is periodic. Hence, it must be possible to express the period T of this motion in terms of the physical parameters, k and m, which characterize the system. Comparison of the solution $x(t)$ in (6-21) with the periodic function $g(t)$ in (6-6) shows that the period T of the motion is $2\pi/\omega$. Hence making use of the definition for ω in (6-19), we find that

$$T = 2\pi \sqrt{\frac{m}{k}} \tag{6-24}$$

Note that T is determined exclusively by the ratio k/m and is in no way related to the initial conditions. Hence, regardless of how the oscillator is

originally displaced from equilibrium, the period of the resultant motion is the same.

Let us now turn to the two parameters A and α, which are determined by the initial conditions. First, since the cosine function is bounded between the limits -1 and $+1$,

$$-1 \leq \cos(\omega t + \alpha) \leq 1$$

for all values of ω, t, and α, it follows that as t varies, the displacement of the particle $x(t)$ in (6-21) ranges between the limits $-A$ to $+A$ regardless of the values for ω and α. Further, x assumes the values $\pm A$ at those times for which $\cos(\omega t + \alpha) = \pm 1$, and for these values of t, the velocity in (6-22) vanishes. Hence, as previously observed, the integration constant A is the amplitude of the motion. If, for example, a particle is displaced from its equilibrium position by a distance h and then released at rest, its position at time t is

$$x(t) = h \cos \omega t$$

for only the choices $A = h$ and $\alpha = 0$ will, when substituted into (6-21), satisfy the stated initial conditions.

To understand the significance of the phase angle α, suppose the particle is initially displaced from its equilibrium position by a distance h and given a velocity v_0 as it is released. In this case, h is *not* the amplitude of the motion nor does the phase angle vanish. To calculate the values of A and α we substitute the given values $x(0) = h$ and $v(0) = v_0$ into (6-21) and (6-22), respectively. The result is

$$h = A \cos \alpha \qquad v_0 = -\omega A \sin \alpha$$

so that

$$\tan \alpha = -\frac{v_0}{\omega h} \qquad A = \sqrt{h^2 + v_0^2/\omega^2} \tag{6-25}$$

Hence, given the initial displacement and velocity of the particle, the phase angle α and the amplitude A are determined uniquely. Only for the special case $v_0 = 0$ is the amplitude the same as the initial displacement.

Consider, for example, the case of a particle initially at its equilibrium position, when it is struck so that it starts to travel at the velocity v_0. In this case the appropriate parameter values are $h = 0$ and $v_0 = v_0$. Substitution into (6-25) yields the values $\alpha = -\pi/2$, $A = v_0/\omega$. Hence, according to (6-21),

$$x(t) = A \cos(\omega t + \alpha) = \frac{v_0}{\omega} \cos\left(\omega t - \frac{\pi}{2}\right) = \frac{v_0}{\omega} \sin \omega t$$

The significance of A and α is further illustrated in Figure 6-10, which shows a plot of $x(t)$ in (6-21) as a function of the variable ωt.

Example 6-7 A 0.2 kg mass attached to the free end of a vibrating spring is observed to have a period of 3.0 seconds.

(a) What is the spring constant?

(b) If its amplitude is 10 cm, what is the maximum velocity?

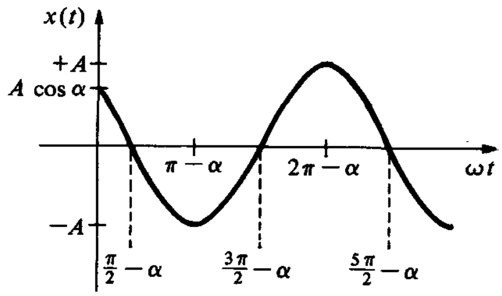


Figure 6-10

Solution

(a) Solving (6-24) for k, we obtain

$$k = \frac{4\pi^2 m}{T^2}$$

and substituting the given parameter values we find that

$$k = \frac{4\pi^2 \times 0.2\ \text{kg}}{(3.0\ \text{s})^2} = 0.88\ \text{N/m}$$

(b) Since $|\sin\theta| \leq 1$ for all θ, it follows from (6-22) that the maximum velocity v_m is ωA. Hence

$$v_m = \omega A = \sqrt{\frac{k}{m}}\,A$$

$$= \sqrt{\frac{0.88\ \text{N/m}}{0.2\ \text{kg}}} \times 0.1\ \text{m} = 0.21\ \text{m/s}$$

Example 6-8 Suppose that a particle having a mass of 10 grams is attached to the lower end of a vertical spring of constant $k = 1$ N/m.

(a) By what amount is the spring stretched from its original length?

(b) What is the value for ω?

(c) What is the position and the velocity of the particle at any time t if it is subsequently displaced upward a distance 3.0 cm and then released at rest?

Solution

(a) The distance d by which the spring is stretched is determined from (6-7) to be

$$d = \frac{mg}{k} = \frac{0.01\ \text{kg} \times 9.8\ \text{m/s}^2}{1.0\ \text{N/m}} = 0.098\ \text{m}$$

(b) According to (6-19), ω is

$$\omega = \sqrt{\frac{k}{m}} = \sqrt{\frac{1\ \text{N/m}}{10^{-2}\ \text{kg}}} = 10\ \text{rad/s}$$

(c) To find $x(t)$, we substitute the initial values $v(0) = 0$, $x(0) = 3.0$ cm into (6-25). The result is

$$\tan\alpha = 0 \qquad A = h = 3.0\ \text{cm}$$

and using these values for A and α, we find by the use of (6-21) and (6-22) that

$$x = 0.03 \cos 10t$$
$$v = -0.3 \sin 10t$$

where t is in seconds, x in meters, and v in meters per second.

Example 6-9 A man of mass 80 kg stands on a horizontal platform which goes up and down with simple harmonic motion of period $T = 4.0$ seconds and of amplitude $y_0 = 20$ cm. Calculate the force F that the platform exerts on him, assuming that at $t = 0$ the platform is at its maximum height.

Solution Define $y(t)$, as shown in Figure 6-11, to be the location at time t of the man's foot relative to the equilibrium position of the platform. Then

$$y = y_0 \cos\left(\frac{2\pi t}{T}\right)$$

since, in this case, $A = y_0$ and $\alpha = 0$. Differentiating twice, we find by use of (6-17) and (6-18) that the acceleration a_y is given by

$$a_y = \frac{d^2y}{dt^2} = -y_0\left(\frac{2\pi}{T}\right)^2 \cos\left(\frac{2\pi t}{T}\right)$$

where the coefficient $y_0(2\pi/T)^2$ has the numerical value 0.49 m/s^2. The net force acting on him is $(F - mg)$ and acts along the y-direction. Applying the second law, we obtain

$$ma_y = F - mg$$

and thus

$$F = m(g + a_y) = 80\ \text{kg}\left[9.8(\text{m/s}^2) - 0.49(\text{m/s}^2)\cos\frac{2\pi t}{T}\right]$$
$$= 780\left(1 - 0.05\cos\frac{2\pi t}{T}\right)\text{N}$$

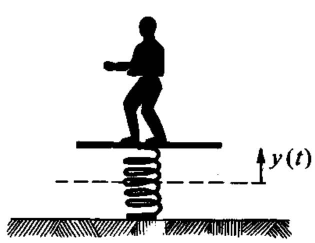


Figure 6-11

6-8 The simple pendulum

Besides the particle oscillating at the end of a spring, there are various other physical systems that can also undergo simple harmonic motion. One of these is the *simple pendulum*. In the present section the motion of this system will be

analyzed, and it will be established that for small amplitudes the simple pendulum executes SHM.

Consider, in Figure 6-12a, a small body of mass m suspended from a fixed point by a massless stick or string of length l. By "small" in the present context is meant a body whose linear dimensions are very small compared to l. If the particle is displaced from its equilibrium position, say by an angle θ_0, and then released, it will travel back and forth along the circular path AB of radius l and centered at the point of suspension. The position of the particle at any time t can, for this type of motion, be completely specified in terms of the angle θ that the string makes with the vertical. Let us adopt the convention that positive values for θ correspond to positions to the right side of the vertical, and negative ones correspond to positions on the left. The equilibrium position occurs at $\theta = 0$. Since the particle travels along a circular path it follows from (5-15) and (5-17) that its velocity v may be expressed in terms of θ by

$$v = l\frac{d\theta}{dt} \tag{6-26}$$

since l is the radius of the orbit in this case.

As shown in Figure 6-12b, the forces acting on the particle are (a) the tension **P** with which the string pulls on it, and (b) the gravitational force $m\mathbf{g}$ acting vertically down. The force $m\mathbf{g}$ can be resolved into a component $mg \cos \theta$ along the string and a component $mg \sin \theta$ along the direction of motion of the particle. According to the analysis in Sections 3-10 and 5-7, when a particle moves along the arc of a circle—as the bob of a pendulum does—it experiences two types of acceleration. One of these is the centripetal acceleration $\mathbf{a}_c$, which, according to (3-25) and Figure 3-23, has the magnitude $l(d\theta/dt)^2$ and is directed radially inward. The second is the tangential acceleration $\mathbf{a}_t$, which, in accordance with (5-14), (5-17), and (6-26), and Figure 3-23, has in the present case the magnitude $l(d^2\theta/dt^2)$ and is directed along a tangent to the circle. Reference to Figure 6-12b shows that the force that

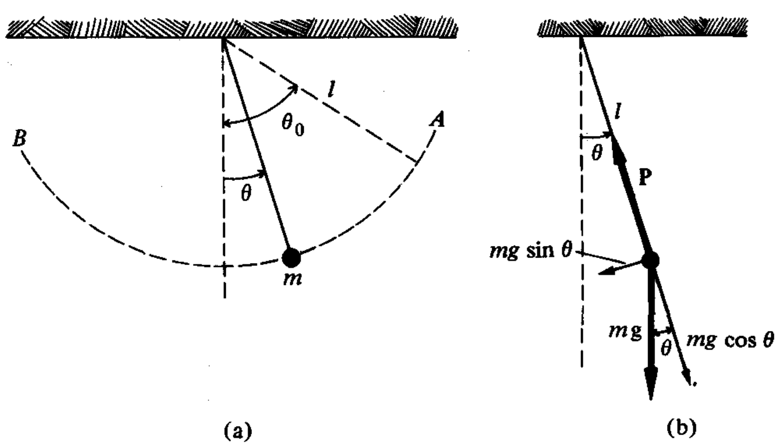


Figure 6-12

produces the centripetal acceleration is $(P - mg \cos \theta)$. Hence an application of the second law yields

$$P - mg \cos \theta = ml \left(\frac{d\theta}{dt}\right)^2 \tag{6-27}$$

Similarly, the component of the force that produces the tangential acceleration is $-mg \sin \theta$, where the minus sign reflects the fact that this component is oriented along the direction of decreasing values for θ. A second application of Newton's second law thus yields

$$ml \frac{d^2\theta}{dt^2} = -mg \sin \theta \tag{6-28}$$

which together with (6-27) determines the angular position θ of the particle and the tension P in the string at any time t.

To simplify matters, let us restrict the following considerations to the case for which the maximum angular displacement θ_0 is small, say $\theta_0 \leqslant 5° \cong 0.1$ rad. It follows (see Appendix C) that, to an accuracy of better than 1 percent, $\theta_0 \cong \sin \theta_0$. Further, assuming the bob is released at rest, it follows that since the angular displacement θ of the particle has a magnitude always less than θ_0, the factor $\sin \theta$ in (6-28) may be approximated simply by θ. Making this replacement and dividing both sides of (6-28) by ml, we obtain

$$\frac{d^2\theta}{dt^2} = -\frac{g}{l}\theta \tag{6-29}$$

Finally, comparison with the equation of motion for the simple harmonic oscillator in (6-20) shows that for small amplitudes the pendulum also executes SHM. The angular frequency ω of this oscillatory motion is

$$\omega = \sqrt{\frac{g}{l}} \tag{6-30}$$

and thus the period is

$$T = 2\pi \sqrt{\frac{l}{g}} \tag{6-31}$$

Assuming that the pendulum initially is displaced an angle θ_0 and released from rest, we find, by comparing with (6-21), that the appropriate solution of (6-29) is

$$\theta(t) = \theta_0 \cos \omega t \tag{6-32}$$

with ω as given in (6-30).

To summarize, then, if a simple pendulum oscillates with small amplitude, its motion is simple harmonic. The equation of motion is given in (6-29), and its solution for the angular displacement $\theta(t)$ is given in (6-32). Note that the period T of the pendulum depends only on its length l and the local value g of

the acceleration of gravity. Thus, by an appropriate choice for l, we can use the pendulum as a device to measure time. Further, a pendulum can also be used to measure local variations in g; this feature has found important applications in geological researches.

Example 6-10 A simple pendulum of length 20 cm undergoes oscillations of small amplitude:

(a) What is the period of the motion?

(b) If it is initially released from rest at an initial displacement of $\theta_0 = 0.1$ rad, what is its maximum velocity?

Solution

(a) Substituting the given value $l = 0.2$ meters into (6-31), we find that

$$T = 2\pi\sqrt{\frac{l}{g}} = 2\pi\sqrt{\frac{0.2\text{ m}}{9.8\text{ m/s}^2}} = 0.90\text{ s}$$

(b) Differentiating (6-32), we find that the maximum value for $d\theta/dt$ is $\omega\theta_0$. Hence,

$$\left(\frac{d\theta}{dt}\right)_{\max} = \omega\theta_0 = \sqrt{\frac{g}{l}}\,\theta_0 = \sqrt{\frac{9.8\text{ m/s}^2}{0.2\text{ m}}}\times 0.1\text{ rad}$$
$$= 0.70\text{ rad/s}$$

Example 6-11 A pendulum undergoes small-amplitude oscillations with a period of 1.5 seconds. If the mass of the bob is 1 kg and the maximum angular displacement is 0.08 rad, calculate:

(a) The length of the pendulum.

(b) The maximum tension in the string.

Solution

(a) Solving (6-31) for l, we find that

$$l = \frac{T^2}{4\pi^2}g$$

and thus

$$l = \frac{(1.5\text{ s})^2}{4\pi^2}\times 9.8\text{ m/s}^2 = 0.56\text{ m}$$

(b) The maximum tension P occurs when the bob is at its lowest point at $\theta = 0$. Why? Here its angular velocity $d\theta/dt$ achieves its maximum $\omega\theta_0$. Hence

$$\left(\frac{d\theta}{dt}\right)_{\max} = \omega\theta_0$$
$$= \frac{2\pi}{1.5\text{ s}}\times 0.08\text{ rad}$$
$$= 0.34\text{ rad/s}$$

Substitution into (6-27) yields for the maximum tension P_m in the supporting string

$$\begin{aligned} P_m &= \left[mg \cos\theta + ml\left(\frac{d\theta}{dt}\right)^2 \right]_{\max} \\ &= m[g + l(\omega\theta_0)^2] \\ &= 1\ \mathrm{kg} \times [9.8\ \mathrm{m/s^2} + 0.56\ \mathrm{m} \times (0.34\ \mathrm{rad/s})^2] \\ &= 9.86\ \mathrm{N} \end{aligned}$$

Thus the tension in the string is not increased very much by the motion of the bob in this case.

†***Example 6-12*** A simple pendulum of length 20 cm is suspended from the ceiling of an elevator. Calculate the period of the pendulum if the elevator has an upward acceleration of 3.0 m/s^2.

Solution Figure 6-13 shows the situation as viewed by an observer at rest in the elevator. The second law is valid here, provided that we add a fictitious force $-m\mathbf{a}_0(a_0 = 3.0\ \mathrm{m/s^2})$, directed vertically downward since the elevator is accelerating upward at $\mathbf{a}_0$. Comparison of this situation with that in Figure 6-12 shows that all of our previous results are valid, provided that g is replaced by $(g + a_0)$. Making use of (6-31), for the period T we obtain

$$\begin{aligned} T &= 2\pi\sqrt{\frac{l}{g + a_0}} = 2\pi\sqrt{\frac{0.2\ \mathrm{m}}{9.8\ \mathrm{m/s^2} + 3.0\ \mathrm{m/s^2}}} \\ &= 0.79\ \mathrm{s} \end{aligned}$$

This is to be compared with the period T_0 in an unaccelerated system

$$T_0 = 2\pi\sqrt{\frac{l}{g}} = 2\pi\sqrt{\frac{0.2\ \mathrm{m}}{9.8\ \mathrm{m/s^2}}} = 0.90\ \mathrm{s}$$

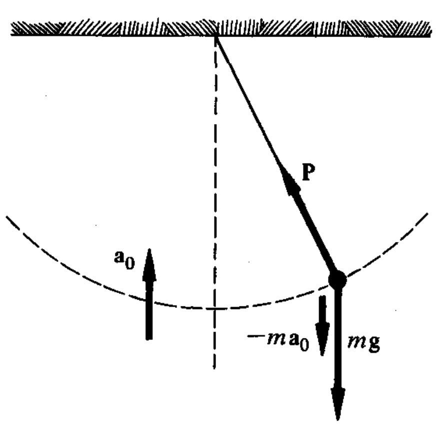


Figure 6-13

6-9 Representation of SHM by circular motion

In addition to the motion of a particle attached to a vibrating spring, and the motion associated with the small-amplitude oscillation of a simple pendulum,

there is another very simple physical system that also exhibits simple harmonic motion. This third system involves uniform circular motion.

Consider, in Figure 6-14, a particle P, which moves counterclockwise with uniform linear velocity v_0 on the rim of a circle centered at O and of radius R. According to (5-15), the angular velocity ω of the particle is v_0/R. If at time $t = 0$ it is at point A, at an angular displacement α with respect to an arbitrary reference line BC, then at time t it will be at a point P, for which the angle AOP has the value ωt. Consider now the motion of point Q, which is the projection of P onto the reference line BC. As P moves around the circle, point Q travels back and forth along the reference line BC. In terms of a coordinate system with origin at O and with x-axis along BC, the x-coordinate of this point Q is

$$x = R \cos(\omega t + \alpha) \tag{6-33}$$

Hence, comparing this formula with (6-21), we see that as the particle moves around the circle with its uniform angular velocity $\omega(= v_0/R)$, its projection Q moves along BC with simple harmonic motion with amplitude equal to the radius R of the circle and with period T given in (6-4). Note that (6-33) describes the motion of Q regardless of how the reference line is selected, provided only that it goes through the center of the circle. Different reference lines correspond to different choices for the phase α.

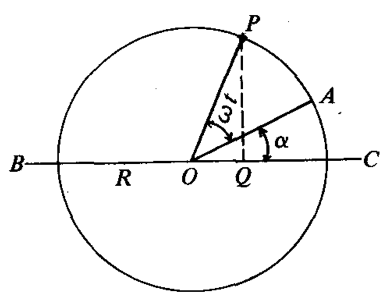


Figure 6-14

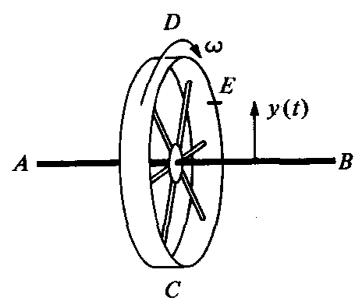


Figure 6-15

Figure 6-15 shows a simple way for producing and observing this motion. Consider a wheel CD, which is suspended on and free to rotate about a horizontal axis AB through its center. Suppose also that there is a marker fixed at the point E on the rim. If the wheel is rotated with uniform angular velocity ω about the axis AB, then the marker at E will travel around a circle and thus undergo periodic motion with period $2\pi/\omega$. The motion of the marker, if viewed in the plane of the wheel and at some distance from it, will appear to consist of an up and down motion with a certain variable velocity v. If the vertical distance of the marker above the axis is described by the coordinate y, then its position at any time t is

$$y = R \cos(\omega t + \alpha)$$

where R is the radius of the wheel and the parameter α is determined by the location of the marker at the initial instant. Thus the observed up and down motion of the marker is SHM.

It is mainly because of this analogy between the motion of a point on a rotating wheel and SHM that the parameter ω, as defined in (6-19), is known as the *angular frequency* of the harmonic oscillator.

†6-10 Damped harmonic motion

On observing simple harmonic motion in the laboratory, we find that as a rule the amplitude of the motion gradually decreases and eventually the oscillating body comes to rest. Thus, although for small angles the motion of a pendulum is SHM, careful observation shows that the maximum angular displacement gradually becomes smaller. The motion in these cases is said to be *damped* and we refer to it as *damped harmonic motion.*

On physical grounds we know that associated with the motion of any macroscopic system there are inevitably frictional effects and in our previous description of SHM, this possibility has been completely ignored. As the bob of the pendulum in Figure 6-12 moves back and forth, for example, in addition to the forces **P** and $m\mathbf{g}$, the bob also experiences a viscous force due to its motion through the air. Similarly, as the block in Figure 6-6 moves back and forth, its amplitude also gradually decreases, partly because of the viscous force of the air and partly because of the frictional forces in the spring itself. In short, simple harmonic motion is strictly an idealization; in practice, the amplitude of this motion gradually becomes smaller because of friction.

To illustrate the physical principle associated with damped harmonic motion, let us see what modifications come about if viscous forces are included in the analysis of a particle attached to a vibrating spring. Assuming for simplicity that this viscous force has the form in (5-18), the equation of motion for the block in Figure 6-6 is

$$m\frac{d^2x}{dt^2} = -kx - bv \tag{6-34}$$

which differs from (6-12) only by the term $-bv$. In this formula, $v = dx/dt$ is the velocity of the block, and the minus sign in front of the viscous-force term reflects the fact that it is always directed so as to oppose motion. If we divide (6-34) through by m, and regroup terms, it may be written

$$\frac{d^2x}{dt^2} + \frac{b}{m}\frac{dx}{dt} + \omega^2 x = 0 \tag{6-35}$$

where $\omega \equiv (k/m)^{1/2}$ is the angular frequency of the corresponding undamped oscillator. Note that the parameters b/m and ω both have the same dimensions of reciprocal time. For the case of present interest, for which the viscous force is very small compared to the force of the spring, these parameters satisfy the inequality

$$\frac{b}{m} \ll \omega \tag{6-36}$$

It is left as an exercise to confirm by use of the product rule for differentiation and the results of Appendix C that a solution of the equation of motion in (6-35) is

$$x(t) = Ae^{-bt/2m} \cos\left[t\sqrt{\omega^2 - \frac{b^2}{4m^2}} + \alpha\right] \tag{6-37}$$

where A and α are arbitrary constants. For the case of no damping, when $b = 0$ (6-37) reduces to the expected form in (6-21). Because of (6-36), the solution in (6-37) may be expressed in a simpler form by replacing the square-root factor $(\omega^2 - b^2/4m^2)^{1/2}$ by ω:

$$x(t) = Ae^{-bt/2m} \cos(\omega t + \alpha) \tag{6-38}$$

Except for the exponential factor $e^{-bt/2m}$, this is the same as the formula for the undamped case. This means that the effect of the weak retarding force, $-bv$, is to cause the particle to undergo a modified simple harmonic motion with an exponentially decaying "amplitude" $Ae^{-bt/2m}$. Figure 6-16 shows a plot of (6-38) for the special choice $\alpha = 0$. Note that the no-damping curve $A \cos \omega t$ is squeezed between the two factors $\pm Ae^{-bt/2m}$.

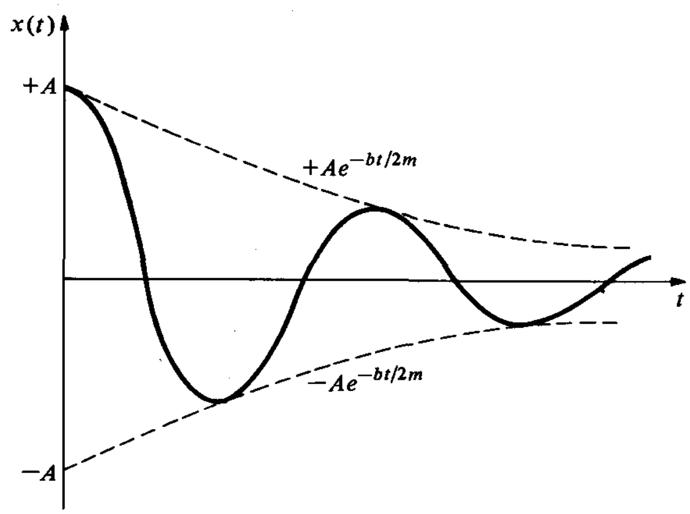


Figure 6-16

†6-11 Forced harmonic motion; resonance

Forced harmonic motion results if a system which would undergo SHM if left to itself, is actually subjected to the action of an additional force of some type. If, for example, the bob of an oscillatory pendulum or a particle vibrating at the end of a spring is struck repeatedly, then forced harmonic motion is produced. A physical system undergoing this type of forced motion is known as a *driven harmonic oscillator.*

Associated with the motion of a driven harmonic oscillator, there is the very

striking phenomenon, known as a *resonance*, which arises if the external driving force is periodic, with a period comparable to the natural period $2\pi(m/k)^{1/2}$ or $2\pi(l/g)^{1/2}$ of the oscillator. In a resonance situation, even though the system may start out with a small amplitude, and even though the strength of the driving force may itself be very small, the amplitude of the motion can become extraordinarily large. If, for example, a particle attached to the end of a spring is repeatedly tapped with a frequency very near the natural or resonant frequency $(k/m)^{1/2}/2\pi$, it is found that the amplitude of the motion steadily increases until finally it is so large that Hooke's law itself ceases to be valid.

Physically, the phenomenon of resonance can be understood in the following way. Suppose you repeatedly tap the bob of a pendulum with a repetition rate ν which is not the same as its natural frequency $(g/l)^{1/2}/2\pi$. Then sometimes during its oscillations you will be tapping it to the right while it is moving left, thereby inhibiting its motion. At other times you will be enhancing its motion when you tap it to the right, while it is traveling to the right. However, if your repetition rate is *not* the same as the natural or resonant frequency of the pendulum, more often than not you will strike it the wrong way, so the amplitude of the oscillations do not tend to become large. On the other hand, if you repeatedly tap the pendulum with a frequency close to its natural frequency $(g/l)^{1/2}/2\pi$, then each time you tap it you will be giving it an additional increment of motion. For this case then, it is easy to see that it is very possible for the pendulum to go into a resonant state involving large-amplitude oscillations.

Figure 6-17 shows an apparatus that may be used to demonstrate this effect. A horizontal rod AB is supported by two strings S_1 and S_2 and has suspended from it three pairs of pendulums, (a, a'), (b, b'), and (c, c'), with the members of each pair having the same length. If we displace one of these pendulums, say c, in a direction perpendicular to the plane of the page, then its resultant oscillatory motion causes in rod AB a very slight perturbing motion whose period is the same as that of c. Because of this motion of the rod AB, each of the remaining pendulums (a, a', b, b', c') experiences a slight periodic force of *period precisely the same as that*

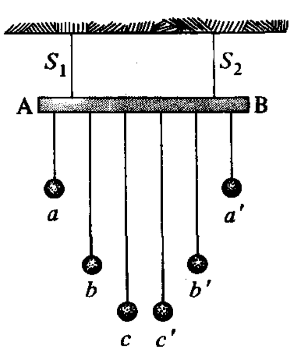


Figure 6-17

of c. Consistent with the preceding discussion, we find that pendulum c', whose length (and therefore whose period) is precisely the same as that of c, starts oscillating back and forth with steadily increasing amplitude. (We shall see in Chapter 8 that its amplitude can never exceed that of c.) On the other hand, the amplitudes of the remaining pendulums remain small throughout the subsequent motion of c and c' since their natural periods are not the same as that of the perturbing force due to rod AB.

Finally, let us consider the concept of forced oscillations and resonance from a quantitative point of view. Consider a particle vibrating at the end of a spring and suppose the existence of an additional force $F_0 \cos \omega_0 t$, with F_0 and ω_0 positive constants. Newton's law of motion has the form

$$m\frac{d^2x}{dt^2} = -kx + F_0 \cos \omega_0 t \tag{6-39}$$

with m the mass of the particle and k the spring constant. Making use of the differentiation formulas in (6-17) and (6-18), it is easy to confirm that a solution of (6-39) is

$$x(t) = A\cos(\omega t + \alpha) - \frac{F_0/m}{\omega_0^2 - k/m}\cos \omega_0 t \tag{6-40}$$

with A and α arbitrary constants, and ω defined in terms of k and m by (6-19). On comparing this with the solution of the nonforced harmonic oscillator in (6-21), we see that the difference is the term

$$\frac{F_0/m}{\omega_0^2 - k/m}\cos \omega_0 t \tag{6-41}$$

which corresponds to oscillatory motion at the *driving* angular frequency ω_0. Hence the particle vibrates with two distinct frequencies: one of these, ω, is determined by the stiffness of the spring in accordance with (6-19), and the second, ω_0, characterizes the angular frequency ω_0 of the driving force. The resulting motion is thus much more complicated.

It is of some interest to consider the solution in (6-40) in its dependency on the angular frequency ω_0 of the driving force. The first term is of no interest here since it is independent of ω_0 and is bounded in magnitude by the constant A. On the other hand, the second term depends in a very sensitive way on ω_0, and for the special case $\omega_0 \approx \omega = (k/m)^{1/2}$, it becomes infinitely large. Thus if ω_0 is equal to the natural angular frequency $(k/m)^{1/2}$ of the oscillator, then the amplitude of the motion can become very large and give rise to a resonance. And consistent with the above discussion, the amplitude becomes very large even though the strength F_0 of the driving force may itself be rather small.

The "infinity" in (6-40) at $\omega_0^2 = k/m$ is, of course, no problem and is a consequence of our assumed idealized conditions. If the effects of friction are included in (6-39), no infinities arise in the solution.

6-12 Summary of important formulas

The equation of motion of a particle of mass m attached to a vibrating spring of constant k is

$$m\frac{d^2x}{dt^2} + kx = 0 \tag{6-12}$$

where x is the displacement of the particle from equilibrium. It has the solution

$$x = A\cos(\omega t + \alpha) \tag{6-21}$$

with A and α integration constants and with ω *defined* in (6-19). The period T of the oscillator is

$$T = 2\pi\sqrt{\frac{m}{k}} \tag{6-24}$$

The equation of motion of a simple pendulum of length l is

$$ml\frac{d^2\theta}{dt^2} = -mg\sin\theta \tag{6-28}$$

and for small angles $\theta\,(\lesssim 5°)$ the motion is simple harmonic with period

$$T = 2\pi\sqrt{\frac{l}{g}} \tag{6-31}$$

QUESTIONS

1. Define or describe briefly what is meant by the following terms: (a) periodic motion; (b) harmonic motion; (c) Hooke's law; (d) resonance; and (e) damped oscillations.
2. Illustrate, by reference to a particular case, a motion for which the particle returns to a fixed point $\mathbf{r}_0$ at the end of each time interval of length 2.0 seconds but for which the period of the motion is *not* 2.0 seconds. (*Hint*: Consider SHM with a period of 4.0 seconds.)
3. Is the solar system an example of a system undergoing periodic motion? If so, what is its period?
4. A particle undergoes periodic motion. Why must the average value for the velocity of the particle over a period vanish? Need the same be true for its acceleration? For its speed? Explain.
5. Why is it that if you suspend one end of a spring from a ceiling, the spring stretches even if no mass is attached to the other end?
6. You observe the period of a particle oscillating at the end of a spring. Can you determine the spring constant from this knowledge? If not, what other datum do you require?
7. If you connect a 10-gram particle to the end of a vertical spring and measure the amount by which the spring is stretched at equilibrium, can you predict the period of oscillation if the system is disturbed subsequently? If not, what other datum do you require?
8. A small body is attached simultane-

ously to two identical springs hanging next to each other. Is the period of the subsequent motion greater or less than it would be if it were attached to only one of the springs?

9. Suppose in Example 6-3 that initially the two springs do not have their natural lengths, but that the block is nevertheless in equilibrium under their combined forces. What would be the force on the block this time if it is displaced a distance d to the right?
10. A particle oscillating at the end of a certain spring is found to have a period of 2.0 seconds. If the spring is cut *in half*, what would be the period of oscillation of the particle if it is vibrating while attached to an end of one of these segments? What would it be if the spring were cut into thirds?
11. Is our assumption that the spring be massless important in the analysis of SHM? Devise an experiment to determine the effect of the spring's mass on the period of a simple harmonic oscillator.
12. At a point where the magnitude of the acceleration of a harmonic oscillator is a maximum, what can you say about the speed of the particle? At a place where its speed is maximum, what can you say about its acceleration?
13. Consider the physical situation in Figure 6-11. Under what circumstances, if any, will the man leave the platform?
14. What type of a curve results if you draw a graph of (6-15) by plotting v as a function of x? What happens to the curve as k increases? As A increases?
15. If you double the length l of a simple pendulum, what happens to its period?
16. Why do you suppose the relation between the period and length of a simple pendulum is not used to define a standard of time in terms of length?
17. At what point along the path of an oscillating pendulum is the tension in the supporting string a maximum? Where is it a minimum?
18. Explain the physical mechanism that causes the amplitude of a pendulum to decrease gradually.
19. What is meant by the term *resonance*? Under what circumstances will a pendulum exhibit a resonance?
20. A particle vibrates on the end of a vertical spring at a certain frequency ν_0. What happens to its frequency if the spring is suspended from the ceiling of an elevator that accelerates upward? What happens if the elevator accelerates downward?

PROBLEMS

1. An oscillating body takes 0.4 second to complete one full cycle of its motion. What are the values for (a) its period? and (b) its frequency?
2. If it takes the planet Jupiter 4333 terrestrial days to complete one orbit about the sun, what are the period, the frequency, and the angular frequency ω of this motion?
3. A wheel rotates about a shaft and completes 1200 revolutions each minute. What is the angular frequency of a point on the rim of the wheel? What is its period?
4. The motion of a particle is described by

$$x = 0.5 \cos\left(\frac{7}{2}\pi t + \frac{\pi}{4}\right)$$

with x in meters and t in seconds. What is the value for (a) the period; (b) the frequency; and (c) the amplitude?
5. For the particle described in Prob-

lem 4 what are the values for **(a)** the maximum displacement; **(b)** the maximum velocity; **(c)** the minimum speed; and **(d)** the maximum acceleration?

6. A particle undergoes SHM according to the formula

$$y = 0.30 \cos\left(\frac{\pi}{2}t + \frac{\pi}{6}\right)$$

where y is in meters and t in seconds. What are the values for **(a)** the initial displacement; **(b)** the initial speed; **(c)** the initial acceleration; and **(d)** the maximum speed?

7. For the oscillator in Problem 6 calculate the values for the displacement, the velocity, and the acceleration at time $t = 1/3$ second.

8. A particle of mass 0.2 kg is found to stretch a certain spring a distance of 3.0 cm. **(a)** What is the spring constant? **(b)** By how much will a 500-gram block stretch the spring?

9. A block of mass 150 grams is attached simultaneously to the lower ends of two identical springs that hang side by side. If, when attached to one of the springs, the block stretches it by a distance of 0.8 cm, by what distance are the two springs stretched at equilibrium?

10. Making use of the chain rule for differentiation, show explicitly that the time derivative of the energy integral in (6-15) yields the equation of motion in (6-11).

11. A particle undergoes simple harmonic motion of amplitude 2.0 cm. By the use of (6-15), calculate the period of the motion if, when the particle travels through the equilibrium position, it has a velocity of 2.0 cm/s.

12. A block of mass 200 grams oscillates at the end of a spring of constant 2.0 N/m. **(a)** What is the period of the motion? **(b)** If its maximum acceleration is 2.0 m/s², what is the amplitude of the motion?

13. A particle of mass 0.4 kg oscillates at the end of a spring with a period of 1.8 seconds.

(a) If its maximum displacement from equilibrium is 20 cm, what is the spring constant?

(b) If initially the particle is at maximum displacement, what is the phase angle?

(c) Determine a formula for the position function $x(t)$.

14. A 1.2-kg block hangs in equilibrium from a spring of constant 3.0 N/m. If it is struck a blow so that initially it starts to travel upward at a speed of 25 cm/s, calculate **(a)** the period of the motion; **(b)** the amplitude of the motion; and **(c)** the phase angle.

15. Suppose that the block in Problem 14 is initially struck a blow so that it starts to travel downward at a speed of 30 cm/s. Write down the position function $x(t)$ and the velocity $v(t)$ for the block at any time t.

16. A 700-gram block attached to a spring is observed to oscillate with a period of 1.5 seconds and an amplitude of 15 cm.

(a) What is the spring constant?

(b) What is the maximum force that the spring exerts on the block?

(c) What is the maximum velocity?

17. A particle of mass 0.03 kg is attached to the lower end of a spring of natural length 60 cm and is then released at rest—that is, with the spring in its relaxed state. If the spring constant k has the value 9.8 N/m, what is the lowest point, measured from the top of the spring, reached by the particle?

18. Repeat Problem 17, but suppose that this time the spring is compressed 10 cm before the particle is released at rest.

19. A particle of mass m is attached to a suspended spring of natural length

l and then released. Show that in terms of a y-axis directed vertically downward and with the origin at the suspension point of the spring, the position of the particle at any time t is

$$y = l + \frac{mg}{k}\left(1 - \cos\sqrt{\frac{k}{m}}\, t\right)$$

20. Repeat Problem 19, but assume this time that the particle is given an original downward velocity v_0.

21. A particle is attached to a spring and given an initial displacement of 12 cm from its equilibrium position. If it is released at rest and if in a time interval of 2.0 seconds it travels a distance of 8.0 cm toward its equilibrium position, calculate **(a)** the amplitude and the phase angle of the resultant motion; **(b)** the period of the motion; and **(c)** the maximum speed of the particle.

22. By making use of the solution for the equation of motion in (6-21), prove the following:

(a) The acceleration is minimum when the particle's displacement is maximum.

(b) The speed is least when the magnitude of the acceleration is maximum.

(c) The acceleration vanishes when the particle is at the equilibrium position.

(d) The speed is maximum when the particle travels through its equilibrium position.

23. A particle executes SHM. Its maximum speed is 10 cm/s, and its maximum acceleration is 25 cm/s^2.

(a) Calculate the amplitude and the period of the motion.

(b) Is the phase angle determined by these data? If not, fix the phase angle and write down a formula for the position of the particle at any time t.

24. A particle executes SHM. If at a distance of 10 cm from its equilibrium position its acceleration is 1.0 m/s^2, what is the angular frequency ω and the period of the motion?

25. The period of oscillation of a 100-gram particle attached to a certain spring is 2.0 seconds. What would be the period if its mass were increased to 150 grams?

26. If the block in Figure 6-5 has a mass m, show that if it is displaced from equilibrium it will oscillate with a period

$$T = 2\pi\sqrt{\frac{m}{k_1 + k_2}}$$

27. A man of mass 75 kg stands on a platform that goes up and down with simple harmonic motion of amplitude 0.5 m (see Figure 6-11). Determine the smallest value for the period of the motion so that the man will not leave the platform. (*Hint*: Explain why the man will leave the platform at an instant when the normal force that the floor exerts on him vanishes.)

28. Derive the equation of motion of the block in Figure 6-6 if the surface is *not* smooth and is characterized by a coefficient of sliding friction μ. Explain the sense in which the added term—compared with (6-12)—varies in time.

29. Make use of the integral

$$\int \frac{dx}{\sqrt{A^2 - x^2}} = -\cos^{-1}\frac{x}{A}$$

and solve the energy integral in (6-15) for $v = dx/dt$ and thus obtain (6-21) directly.

30. A pendulum consists of a particle of mass 100 grams suspended from a string of length 1.5 meters.

(a) What is the period of oscillation?

(b) If initially the particle is displaced from its equilibrium posi-

tion by 3°($\cong$0.052 rad), what is its angular position at any subsequent time t?

31. A simple pendulum of length 2 meters hangs vertically in equilibrium. The bob is then struck in such a way that it starts to travel horizontally at an initial speed of 2.0 cm/s.
    (a) What is the initial angular velocity?
    (b) What is the period of the pendulum?
    (c) What is the maximum angular displacement?

32. The Foucault pendulum in the Chicago Museum of Science and Industry is observed to have a period of approximately 10 seconds while it undergoes oscillations of small amplitude. What is the length of the pendulum? (*Note*: The Foucault pendulum is used to demonstrate the fact that the earth rotates about its axis. Experiment shows that, at the north or the south pole of the earth, for example, the plane of the oscillating pendulum makes one complete rotation about the vertical during each sidereal day.)

33. Show that if the length l of a simple pendulum is increased very slightly by an amount dl, then the associated fractional change in its period dT/T is

$$\frac{dT}{T}=\frac{1}{2}\frac{dl}{l}$$

34. A simple pendulum of mass m and length l hangs at rest.
    (a) What is the tension in the supporting string?
    (b) If the bob is struck a blow so that it starts to travel horizontally with a speed v_0, calculate the tension in the supporting string immediately afterward.
    (c) Explain in physical terms the difference in your answers to (a) and (b).

35. Show by differentiation that the equation of motion for the simple pendulum in (6-28) admits the energy integral

$$\frac{1}{2}ml\left(\frac{d\theta}{dt}\right)^2-mg\cos\theta = -mg\cos\theta_0$$

where θ_0 is the maximum angular displacement.

†36. A particle of mass m is attached to the ceiling of a moving train by a massless string of length l. At an instant when the train has an acceleration $\mathbf{a}_0$:
    (a) Show that at equilibrium the string will make an angle γ (Figure 6-18) with the vertical, given by

$$\gamma=\tan^{-1}\frac{a_0}{g}$$

    (b) What is the tension in the string at equilibrium?
    (c) If the particle is slightly displaced from equilibrium calculate the resultant period of oscillation.

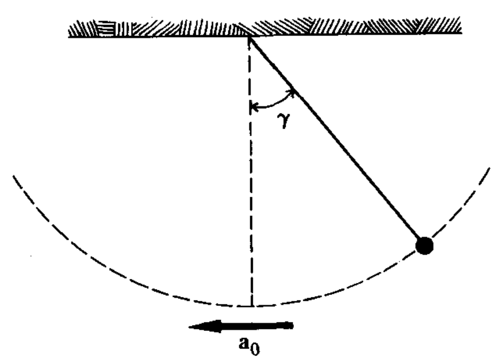


Figure 6-18

†37. What would be the period of the pendulum in Figure 6-13 if the elevator had a downward acceleration a_0? What happens to your answer if $a_0>g$?

†38. A particle of mass m hangs, in equilibrium, from a spring of constant k at a distance l below the ceiling of an elevator.

(a) If the elevator starts to accelerate upward with an acceleration $\mathbf{a}_0$, how far below the ceiling will be the new equilibrium position?

(b) Repeat (a) if the acceleration $\mathbf{a}_0$ is downward.

(c) Suppose that the particle is disturbed from its equilibrium position in either (a) or (b). Show that the period of motion is the same as if the elevator were not accelerating.

†39. A particle of mass 10 grams is suspended by a spring of constant 10 N/m from the ceiling of an elevator. Suppose that initially the system is at rest. The elevator then suddenly starts to accelerate upward at 100 cm/s^2.

(a) Describe the position of the particle at any subsequent time as seen by an observer at rest in the elevator.

(b) Describe the trajectory of the particle as seen by an observer in a Newtonian frame.

40. A *Lissajous figure* is the plane curve traced out by a particle subject to the simultaneous action of two springs at right angles to each other. If ω_1 and ω_2 are the values of the angular frequencies associated with the springs, then the position of the particle at any time may be expressed

$$x = A_1 \cos(\omega_1 t + \alpha_1)$$
$$y = A_2 \cos(\omega_2 t + \alpha_2)$$

with A_1, A_2, α_1, and α_2 constants of integration. Show that if $\omega_1 = \omega_2$, then the Lissajous figure is an ellipse. Show also that if ω_1/ω_2 is a rational number, then the figure is closed.

7 Work and kinetic energy

... To find then what remains unaltered in the phenomena of nature, to discover the elements thereof and the mode of their interconnection and interdependence, this is the business of physical science.

ERNST MACH (1838–1916)

7-1 Introduction

In Chapters 5 and 6 the motion of a particle subject to the action of various forces was analyzed. For most of the cases considered there, we succeeded in obtaining explicit solutions of the Newtonian equations of motion and thus were in a position to predict in all detail the future positions of the particle at any time t.

In addition to these soluble systems, frequently physical situations arise for which explicit solutions of the equations of motion cannot be obtained so readily; for example, a pendulum undergoing large-amplitude oscillations. For these cases, it is sometimes possible, nevertheless, to make predictions of the future behavior of the system in terms of a certain functional relation involving only the position and the velocity of the particle, but not its acceleration. In our studies of motion in the inverse square force field, for example, it was established that the quantity $(\frac{1}{2}mv^2 - GMm/y)$ is constant in time. On equating this quantity to its initial value, we obtained the relation (5-11) between the velocity v and the position y of the moving particle. In

this way, although it was not possible to predict in all detail the particle's trajectory, we were able nevertheless to deduce many qualitative and some quantitative features of its motion.

A quantity of this type, which depends only on the position and the velocity of a particle but which itself does not change in time, is called a *constant of the motion.* For a particle moving in an inverse-square force field, for example, since the quantity $(\frac{1}{2}mv^2 - GMm/y)$ is constant in time, it is also a constant of the motion. Similarly, for the simple harmonic oscillator, the quantity $(\frac{1}{2}mv^2 + \frac{1}{2}kx^2)$ is constant in time according to (6-15), and thus it is also a constant of the motion. These two cases, as we shall see, are representative of a certain class of constants of the motion known as the *energy.* The fact that the energy of certain systems does not change in time is also known as the *law of conservation of energy.*

The main purpose of this chapter and the next is to study this notion of energy and its conservation. We begin this discussion in this chapter by considering the ideas of work and of kinetic energy and the very important work-energy theorem which relates them. Then in Chapter 8 we turn to a development of the law of energy conservation. Although only the case of a single particle will be considered here, it should be kept in mind that the basic ideas carry over to systems of more than one particle. Moreover, many of these ideas are also applicable to situations for which the laws of Newtonian mechanics are superseded by those of the theories of relativity and of quantum mechanics. Energy, its conservation, and the various concepts related to it play an important—and often dominating—role not only in mechanics but in the various branches of engineering, and the chemical and physical sciences as well. *Energy* is truly one of the most fruitful and important physical concepts discovered by man.

7-2 The simple pendulum

Before going into a detailed discussion of the notions of work and kinetic energy, in this section we shall illustrate the type of information that can be obtained by use of a conservation law, by reference to the simple pendulum.

Consider in Figure 7-1 a simple pendulum and suppose that its bob is initially displaced by an angle θ_0 to the point A and then released. Galileo was the first to observe and to record the fact that the maximum angular displacement that the bob achieves on the other side of the vertical is precisely this same angle θ_0. That is, he found that if the bob is released at rest from the point A, a certain vertical distance h below the point of suspension, then it will reach the point B, which is on the opposite side of the vertical and also a distance h below the ceiling. Thus, even though it was not possible for him to predict the motion in all detail, he could predict that no matter at what distance below the ceiling he released the bob, its maximum elevation throughout its subsequent oscillations was precisely the same as

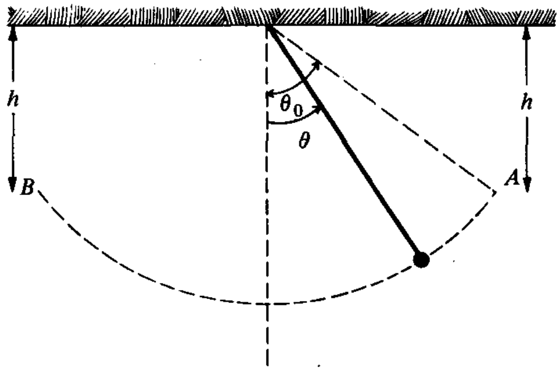


Figure 7-1

the initial value. As we shall see, this constancy of the maximum elevation of the bob of the pendulum is typical of the predictions that can be easily made by use of the law of energy conservation.

Remarkably enough, Galileo carried these observations on the motion of a pendulum even further. Consider the experiment in Figure 7-2, which again portrays a pendulum, but this time with a small peg at the point C just below the point of suspension. As the bob is released from the point A, this time it descends along the circular arc AB, and Figure 7-2b shows the situation at the subsequent instant when the string is vertical. Here the bob is at the lowest point D in its orbit. Now if the peg at C were not present, this would be a repetition of the previous situation and the bob would continue its journey to the point B, defined so that $\angle DOB = \theta_0$. However, the peg *is* present. Thus the motion of that portion of the string labeled OC in Figure 7-2c is arrested while the remainder continues on. This means that the bob of the pendulum now travels along the arc DE of a circle of radius CD centered at the peg. The key point of this experiment is that the final height to which the bob rises at the point E is precisely the same[1] as the initial height at point A. That is, the initial distance h below the ceiling at the point A, is precisely the same as the final distance below the ceiling of the point E.

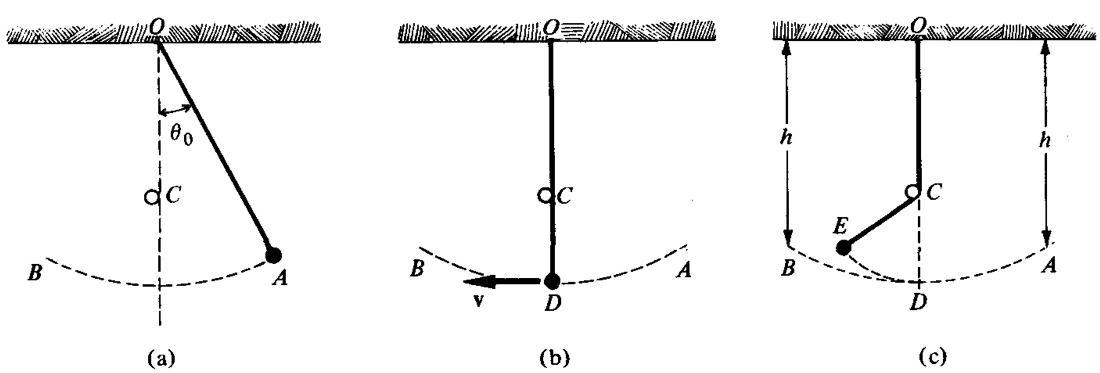


Figure 7-2

[1]That is, assuming that the peg is high enough so that $\overline{OC} - \overline{CE} < h$. Otherwise the string will wrap itself around the peg.

Moreover, experiment shows that regardless of where along the vertical the peg is placed and regardless of the initial angular displacement θ_0, the bob will always achieve a maximum elevation equal to its initial value.

This experiment can be made even more complex by placing several pegs C_1 and C_2 at various places, as in Figure 7-3. If initially the bob is released at rest at the point A, then because of the existence of the pegs at C_1 and C_2, various parts of the string will become arrested until ultimately the bob winds up at the point B. Again, careful measurement shows that the final position of the bob is at the same distance h below the ceiling as was its initial position. As we shall see later, these features are very easy to establish by use of the conservation-of-energy principle, even though the detailed motion of the bob is itself very complex.

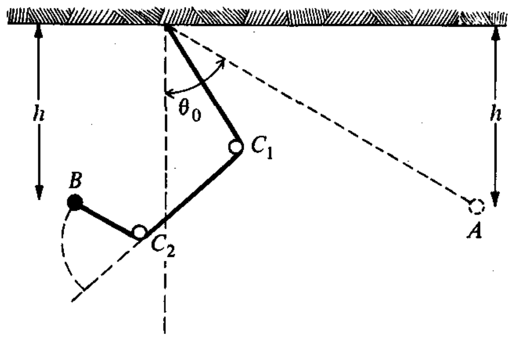


Figure 7-3

7-3 Work

One of the very important physical quantities associated with the concept of energy is that of *work*. In everyday usage, by the term "work" we generally mean an activity of some type that involves physical or mental effort and whose goal is the achievement of some definite, well-defined objective. By contrast, in scientific work this term has a much more restricted meaning.

Consider, in Figure 7-4, a body that is acted upon by various forces which cause it to undergo a displacement **D**. The figure shows only one of these forces $\mathbf{F}_0$, which is singled out for special attention and which, for the moment, we suppose to be a constant force. We define the *work W* carried out by the constant force $\mathbf{F}_0$ when the body undergoes the displacement **D** to

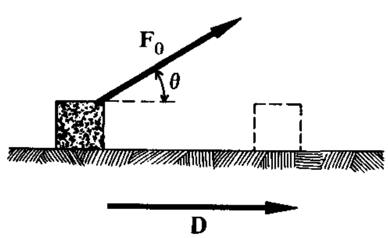


Figure 7-4

be the product of the magnitude D of $\mathbf{D}$ and the component of $\mathbf{F}_0$ along $\mathbf{D}$. If, as shown in the figure, θ is the angle between $\mathbf{F}_0$ and $\mathbf{D}$, then according to (3-6), the component of $\mathbf{F}_0$ along $\mathbf{D}$ is $F_0 \cos\theta$. Hence the work W carried out by the force $\mathbf{F}_0$ is

$$W = F_0 D \cos\theta \tag{7-1}$$

For the special case in which $\mathbf{F}_0$ acts along $\mathbf{D}$, then $\theta = 0$, and here (7-1) becomes

$$W = F_0 D \qquad (\mathbf{F}_0, \mathbf{D} \text{ parallel})$$

Similarly, for the case that $\theta = \pi/2$, which corresponds to the force being perpendicular to the displacement, the work W is zero, since $\cos \pi/2 = 0$. Finally, if $\mathbf{F}_0$ and $\mathbf{D}$ are antiparallel, then $\theta = \pi$ and, since $\cos\pi = -1$, it follows that the W has the negative value $-F_0 D$.

If an agent carries out work W in displacing a body, then we shall also say that during this process the body carries out work $-W$ on the agent. Thus if you carry out work W_0 in pushing a table across a room, then the table carries out work on you in the amount of $-W_0$. Negative work is always to be understood in this sense.

It is important to bear in mind that the force $\mathbf{F}_0$ that appears in the definition of work in (7-1) is not necessarily the only force acting on the body. In general, various other forces will be acting on the body as it is displaced, and when speaking of "work" we have in mind a particular one of these. If a man raises a body of mass m through a vertical height h, for example, then there are two forces acting on the body; one is a downward force mg due to gravity and the second is the compensating upward push of the man. According to (7-1), the work W_M carried out by the man is mgh. On the other hand, since the force of gravity is downward, a second application of (7-1) shows that the work W_G carried out by the earth's gravitational field as the body is displaced vertically upward is $-mgh$. Similarly, if a man lowers the body through a vertical distance h, then the work he carries out is $-mgh$, whereas the work carried out by the gravitational field of the earth is $+mgh$ in this case. These features are illustrated in Figure 7-5.

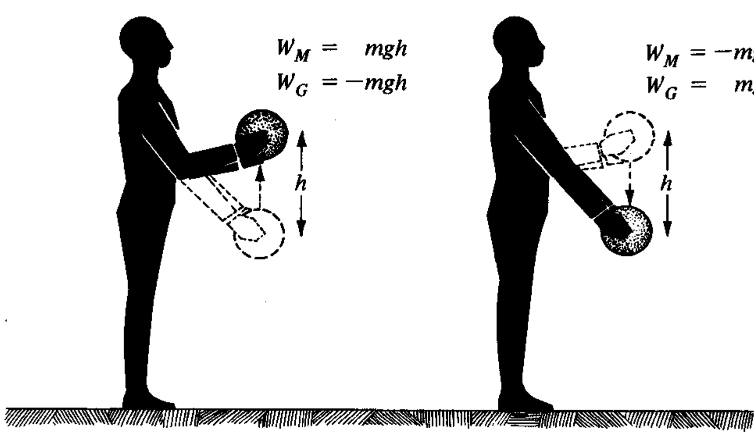


Figure 7-5

It follows from the definition in (7-1) that the SI unit of work is the newton meter. Let us define the unit of work called the *joule* (J) to be 1 newton meter (N-m):

$$1\,\mathrm{J} = 1\,\mathrm{N\text{-}m} \tag{7-2}$$

Thus, if you exert an upward force of 10 newtons to raise a body through a vertical displacement of 2 meters, then, in terms of (7-1), $F_0 = 10$ newtons, $\cos\theta = 1$, and $D = 2$ meters. Hence

$$W = F_0 D \cos\theta = 10\,\mathrm{N} \times 1 \times 2\,\mathrm{m} = 20\,\mathrm{J}$$

Other units of work are the *erg* and the *foot-pound.* These are defined in terms of the joule by the relations

$$1\,\mathrm{J} = 10^7\,\mathrm{erg} = 0.7376\,\mathrm{ft\text{-}lb} \tag{7-3}$$

In studies involving microscopic matter, a frequently used unit is the electron volt (eV). It is defined to be 1.60219×10^{-19} joule.

Example 7-1 A man raises a 60-kg sack a distance of 1.5 meters to his shoulder, carries it for a distance of 300 meters along a level road, and then deposits it on the ground. How much work does he do?

Solution In raising the sack he carries out positive work W, given by

$$\begin{aligned} W = F_0 D &= (60\,\mathrm{kg} \times 9.8\,\mathrm{m/s^2}) \times 1.5\,\mathrm{m} \\ &= 880\,\mathrm{J} \end{aligned}$$

While transporting it along the horizontal road he does no work, since the force he exerts is upward and thus is perpendicular to the displacement. Finally, when he lowers the sack he carries out negative work given by

$$\begin{aligned} W = F_0 D \cos\pi &= (60\,\mathrm{kg} \times 9.8\,\mathrm{m/s^2}) \times (1.5\,\mathrm{m}) \times (-1) \\ &= -880\,\mathrm{J} \end{aligned}$$

Thus it follows that the total work carried out is

$$880\,\mathrm{J} + 0\,\mathrm{J} - 880\,\mathrm{J} = 0$$

Example 7-2 A block of mass m is dragged a distance l up a smooth, inclined plane of angle α by a man who exerts a constant force $\mathbf{F}_0$ parallel to the plane and in a way so that the block has the constant acceleration $\mathbf{a}_0$ (see Figure 7-6).

(a) What force does the man exert?
(b) How much work does he do?
(c) How much work is carried out by the gravitational field of the earth?

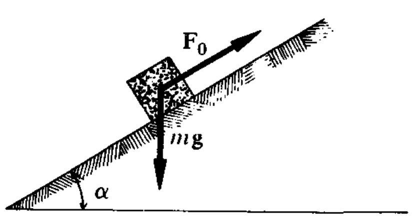


Figure 7-6

Solution

(a) Since the component of the weight mg acting down the plane is $mg \sin \alpha$, it follows from Newton's second law that

$$F_0 - mg \sin \alpha = ma_0$$

whence

$$F_0 = m(a_0 + g \sin \alpha)$$

(b) Substitution into (7-1) yields

$$W = F_0 D \cos \theta = F_0 l = ml(a_0 + g \sin \alpha)$$

(c) The gravitational force mg can be split into a component $mg \cos \alpha$ perpendicular to the plane and a component $mg \sin \alpha$ along and down the plane. The former does no work in displacing the block along the plane while the work W_E carried out by the latter is, according to (7-1),

$$W_E = (-mg \sin \alpha)l$$

The minus sign here reflects the fact that the displacement is directed *upward* along the plane, whereas the component of mg along the plane is directed downward.

7-4 The dot product

The product in (7-1) of the magnitude of a vector and the component of a second vector along the first comes up sufficiently often that it is convenient to designate it in a special way. The purpose of this section is to describe a recipe for multiplying two vectors, which is known as the *dot product* or the *scalar product.*

Consider, in Figure 7-7, two vectors **A** and **B** and let θ be the angle between them. The *dot* or the *scalar* product of **A** and **B** is defined to be the product of the magnitudes of the two vectors and the cosine of the angle between them. Introducing the notation $\mathbf{A} \cdot \mathbf{B}$ for the dot product, we have, by definition,

$$\mathbf{A} \cdot \mathbf{B} = AB \cos \theta \tag{7-4}$$

Reference to the figure shows that $\mathbf{A} \cdot \mathbf{B}$ can be thought of as the product of $|\mathbf{A}|$ and the component $B \cos \theta$ of **B** along **A**. Equivalently, since $A \cos \theta$ is also the component of **A** along **B**, $\mathbf{A} \cdot \mathbf{B}$ can also be viewed as the product of $|\mathbf{B}|$ and the component of **A** along **B**. It follows that the dot product is a

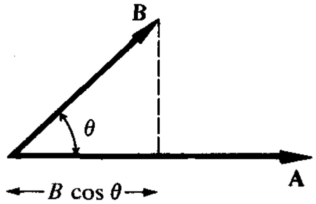


Figure 7-7

commutative relation in that, for any two vectors,

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} \tag{7-5}$$

A second very important property of the scalar product of two vectors is the *distributive law.* This states that for any three vectors **A**, **B**, and **C**

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C} \tag{7-6}$$

In words, this says that the dot product of **A** with the vector $(\mathbf{B}+\mathbf{C})$ is the same as the sum of the dot products of **A** with **B** and of **A** with **C**. Figure 7-8 contains a geometric proof for this distributive law. It is evident from the construction there that the sum of the components of **B** and **C** along **A** is precisely the same as is the component of $(\mathbf{B}+\mathbf{C})$ along the same direction. In terms of the angles θ_B, θ_C, and θ between **A** and the vectors **B**, **C**, and $(\mathbf{B}+\mathbf{C})$, respectively, this may be expressed as

$$B \cos \theta_B + C \cos \theta_C = |\mathbf{B} + \mathbf{C}| \cos \theta$$

and if we multiply throughout by A, (7-6) results.

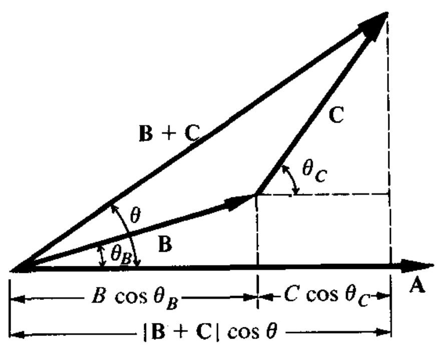


Figure 7-8

Let us express the definition in (7-1) of the work carried out by a constant force in terms of this dot-product notation. Comparison of (7-1) and (7-4) shows that

$$W = \mathbf{F}_0 \cdot \mathbf{D} \tag{7-7}$$

and from now on we shall take this as our definition of work carried out by a constant force.

As an illustration of the dot product, consider the special case $\mathbf{A} = \mathbf{B}$. It follows from (7-4) that, since $\cos 0^\circ = 1$,

$$A = (\mathbf{A} \cdot \mathbf{A})^{1/2} \tag{7-8}$$

and thus the magnitude of a vector is expressible directly in terms of the scalar product of that vector with itself. In particular, if **i** and **j** are basis vectors in a certain coordinate system, then

$$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = 1 \tag{7-9}$$

Moreover, since **i** and **j** are mutually perpendicular, it also follows from (7-4) that

$$\mathbf{i} \cdot \mathbf{j} = 0 \tag{7-10}$$

Example 7-3 Two vectors **A** and **B** have the respective components A_x, A_y and B_x, B_y in a given coordinate system. Calculate $|\mathbf{A}|$ and $|\mathbf{B}|$ and $\mathbf{A} \cdot \mathbf{B}$ in terms of these components.

Solution To calculate $|\mathbf{A}|$, let us make use of the distributive law, (7-6), and the fact that $\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y$. Whence

$$\begin{aligned}\mathbf{A} \cdot \mathbf{A} &= (\mathbf{i}A_x + \mathbf{j}A_y) \cdot (\mathbf{i}A_x + \mathbf{j}A_y) = \mathbf{i} \cdot \mathbf{i}A_x^2 + \mathbf{j} \cdot \mathbf{j}A_y^2 + 2\mathbf{i} \cdot \mathbf{j}A_xA_y \\ &= A_x^2 + A_y^2\end{aligned}$$

where the last equality follows from (7-9) and (7-10). Making use of (7-8) we can conclude, in agreement with (3-8), that

$$A = (A_x^2 + A_y^2)^{1/2}$$

A similar formula holds for B.

In the same way we find that

$$\begin{aligned}\mathbf{A} \cdot \mathbf{B} &= (\mathbf{i}A_x + \mathbf{j}A_y) \cdot (\mathbf{i}B_x + \mathbf{j}B_y) \\ &= \mathbf{i} \cdot \mathbf{i}A_xB_x + \mathbf{j} \cdot \mathbf{j}A_yB_y + \mathbf{i} \cdot \mathbf{j}(A_xB_y + A_yB_x) \\ &= A_xB_x + A_yB_y\end{aligned}$$

where, again, (7-9) and (7-10) have been used.

Example 7-4 A body undergoes a displacement $\mathbf{D} = \mathbf{i}\alpha - \beta\mathbf{j}$ under the action of a constant force $\mathbf{F}_0 = \mathbf{i}a + \mathbf{j}b$, where **i** and **j** are unit vectors along the axes of a given coordinate system. Assuming $\alpha = \beta = 2.0$ meters, $a = 3.0$ newtons, and $b = 4.0$ newtons, calculate:

(a) The magnitudes of $\mathbf{F}_0$ and **D**.

(b) The work carried out by the force $\mathbf{F}_0$.

Solution

(a) According to (3-8), the magnitude of a vector with components A_x and A_y is $(A_x^2 + A_y^2)^{1/2}$. Hence

$$F_0 = (a^2 + b^2)^{1/2} = (3.0^2 + 4.0^2)^{1/2}\ \mathrm{N} = 5.0\ \mathrm{N}$$

and

$$D = (\alpha^2 + \beta^2)^{1/2} = (2.0^2 + 2.0^2)^{1/2}\ \mathrm{m} = 2.8\ \mathrm{m}$$

(b) Substituting the given forms for F_0 and D into (7-7), we obtain

$$\begin{aligned}W = \mathbf{F}_0 \cdot \mathbf{D} &= (\mathbf{i}a + \mathbf{j}b) \cdot (\mathbf{i}\alpha - \mathbf{j}\beta) = a\alpha\mathbf{i} \cdot \mathbf{i} - b\beta\mathbf{j} \cdot \mathbf{j} + \mathbf{i} \cdot \mathbf{j}(\alpha b - a\beta) \\ &= a\alpha - b\beta = 3.0\ \mathrm{N} \times 2.0\ \mathrm{m} - 4.0\ \mathrm{N} \times 2.0\ \mathrm{m} \\ &= -2.0\ \mathrm{J}\end{aligned}$$

The third equality follows by the use of (7-6), and the fourth by the use of (7-9) and (7-10). The fact that $W < 0$ implies that the component of $\mathbf{F}_0$ along **D** is negative.

Example 7-5 Calculate the angle θ between the two vectors $\mathbf{F}_0$ and $\mathbf{D}$ of Example 7-4.

Solution Solving (7-4) for $\cos\theta$ we obtain

$$\cos\theta = \frac{\mathbf{F}_0 \cdot \mathbf{D}}{F_0 D}$$

and, substituting for F_0, D, and $\mathbf{F}_0 \cdot \mathbf{D} = W$ from Example 7-4, we obtain

$$\cos\theta = \frac{\mathbf{F}_0 \cdot \mathbf{D}}{F_0 D} = \frac{-2.0\ \mathrm{J}}{5.0\ \mathrm{N} \times 2.8\ \mathrm{m}} = -0.14$$

and this corresponds to an angle

$$\theta = 98°$$

7-5 Nonconstant forces

The concepts of work and energy are found to be useful mainly for those situations for which the forces acting on the system are either constant or else vary with position. The purpose of this section is to generalize the considerations of Section 7-3 to the case of a force that varies with position in space. For reasons of simplicity, it will be assumed that the body moves along a straight line, the x-axis, and that $F(x)$ is the component of $\mathbf{F}$ along this direction.

To introduce the notion of work carried out by such a variable force, let us consider again briefly the case of a constant force F_0, which moreover acts along the direction of the displacement of a body. Figure 7-9 shows a plot of such a force $F(x)$ as a function of the position x of the body. Since $F(x)$ is a constant, the graph is a horizontal line. Assuming that the body is displaced from the point x to the point $(x + d)$, we find by the use of (7-7) that the work W is the product $F_0 d$. Thus the shaded region in the figure, which is the area under the $F(x)$ curve between the vertical lines erected at x and at $(x + d)$, is a measure of this work.

Let us now turn to the more general case, in which the force $F(x)$ varies explicitly with the position coordinate x of the body. Suppose such a force displaces a small body an infinitesimal distance Δx from the point x to the neighboring point $(x + \Delta x)$. The infinitesimal work ΔW carried by this force is

$$\Delta W = F\,\Delta x \tag{7-11}$$

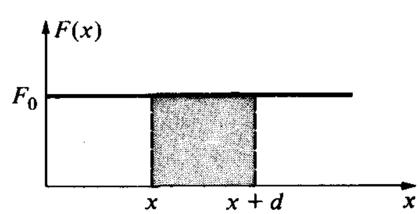


Figure 7-9

where $F = F(x)$ is the value of the force at the point x, and it has been assumed that Δx is sufficiently small that F does not vary appreciably over the interval. Figure 7-10 shows a typical curve $F(x)$. According to (7-11), the work ΔW may be represented on this plot by the shaded area in the figure. Note that the smaller the interval Δx, the more negligible will be the unshaded triangular part below the curve and just above the shaded area, and the more nearly will ΔW be represented by the full area under the curve for $F(x)$.

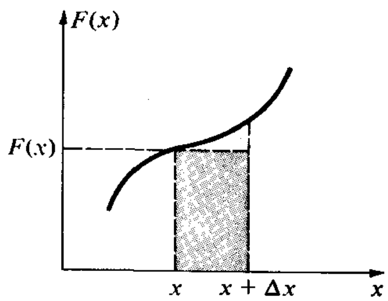


Figure 7-10

By repeated application of the definition in (7-11), we may calculate the work $\overline{W}_{12}$ carried out by a variable force $F(x)$ when it displaces a particle over a finite interval from an initial point, say x_1, to a final point x_2. Imagine first moving the body from the initial point x_1 to the nearby point $(x_1 + \Delta x)$, then moving it from $(x_1 + \Delta x)$ to $(x_1 + 2\Delta x)$, and proceeding this way until ultimately the body is at x_2. The work $\overline{W}_{12}$ carried out by this variable force in displacing the body in this way from x_1 to x_2 is

$$\overline{W}_{12} = F(x_1)\Delta x + F(x_1 + \Delta x)\Delta x + F(x_1 + 2\Delta x)\Delta x + \cdots + F(x_2 - \Delta x)\Delta x \tag{7-12}$$

where there are $(x_2 - x_1)/\Delta x$ terms, one for each interval. Now just as the work ΔW in (7-11) is represented by the shaded area in Figure 7-10, in the same way the work $\overline{W}_{12}$ in (7-12) can be represented (assuming six intervals) by the shaded area in Figure 7-11a. Note that, *in addition to some unshaded*

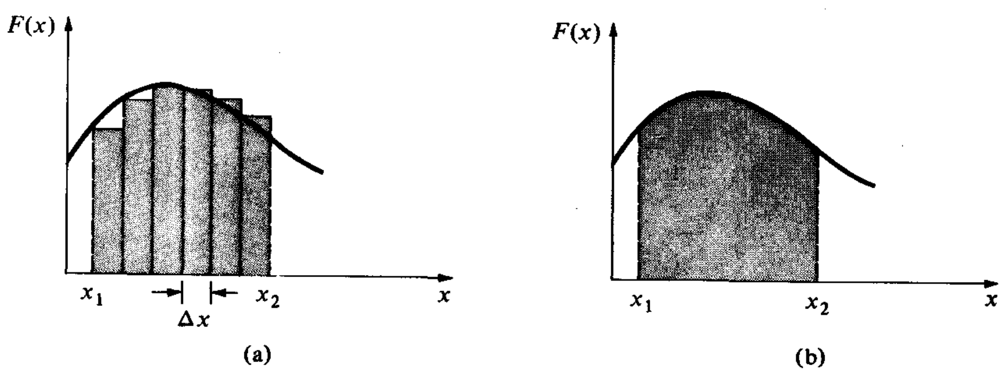


Figure 7-11

triangular areas below the $F(x)$ curve, there are also shaded triangular areas above this curve whenever it has a negative slope.

The purpose of the bar over the symbol $\overline{W}_{12}$ in (7-12) is to emphasize the fact that the work calculated in accordance with this formula depends explicitly on the length of the interval Δx and that, in general, $\overline{W}_{12}$ will assume different values for different choices for Δx. Because of this ambiguity it is convenient to introduce a new quantity W, which is defined to be the limit of $\overline{W}_{12}$ as the length of the interval Δx tends to 0. By definition, then,

$$W = \lim_{\Delta x \to 0} \overline{W}_{12} \tag{7-13}$$

and W may be represented, as in Figure 7-11b, by the full area under the curve $F(x)$ enclosed between the two vertical lines at x_1 and x_2. From now on, by the term "work" carried out by a force we shall always have in mind the quantity W as defined in (7-12), (7-13).

From a practical point of view, it might appear at first to be a very difficult task to calculate the work W associated with any given force field $F(x)$ by use of (7-13). For to use this formula, it appears to be necessary to evaluate $\overline{W}_{12}$ in (7-12) for a sequence of values for Δx tending to 0 and then obtain W as the limit of the associated sequence of values for $\overline{W}_{12}$. Fortunately, this calculation need never be carried out. For according to a fundamental theorem of the calculus, the limit W as defined in (7-12) and (7-13) may be expressed as the difference

$$W = U(x_2) - U(x_1) \tag{7-14}$$

where the derivative of $U(x)$ is the force function $F(x)$; that is,

$$\frac{dU}{dx} = F(x) \tag{7-15}$$

This means that in order to calculate the work W associated with a given $F(x)$, it is necessary only to find a function $U(x)$ whose derivative is $F(x)$. Since the derivative of a constant is zero, it is evident that associated with any given form for $F(x)$ there are, in general, infinitely many functions $\{U(x)\}$, each of which differs from the others by a constant. The work W in (7-13) is given by the difference in (7-14) for *any one* of these functions $U(x)$.

More formally, the result in (7-14) and (7-15) is usually expressed as follows: According to a fundamental theorem of the integral calculus, the shaded area under the curve in Figure 7-11b is the limit of the sum in (7-12), and defines the *definite integral* of $F(x)$ between the limits x_1 and x_2. Equating this definite integral to W, we have

$$W = \int_{x_1}^{x_2} F(x)\, dx \tag{7-16}$$

and (7-14) and (7-15) follow directly by the use of certain properties of the definite integral. In terms of the function $U(x)$ defined in (7-15), the definite integral in (7-16) may be expressed as

$$\int_{x_1}^{x_2} F(x)\,dx = U(x)\Big|_{x_1}^{x_2} = U(x_2) - U(x_1)$$

Example 7-6 Calculate the work carried out by a force $F(x) = \alpha x^2 - \beta$ in displacing a body from x_1 to x_2. Assume the values $\alpha = 3.0\ \mathrm{N/m^2}$, $\beta = 2.0$ newtons, $x_1 = 1.0$ meter and $x_2 = 3.0$ meters.

Solution Using the fact that the derivative of $[(\alpha x^3/3) - \beta x]$ is $(\alpha x^2 - \beta)$ and employing (7-14) through (7-16), we obtain

$$W = \int_{x_1}^{x_2} (\alpha x^2 - \beta)\,dx = \left(\frac{\alpha x^3}{3} - \beta x\right)\Big|_{x_1}^{x_2} = \left(\frac{\alpha x_2^3}{3} - \beta x_2\right) - \left(\frac{\alpha x_1^3}{3} - \beta x_1\right)$$

where, as defined above, the notation

$$U(x)\Big|_{x_1}^{x_2}$$

means the difference $U(x_2) - U(x_1)$. Substituting the given parameter values, we find that

$$W = 22\ \mathrm{J}$$

Example 7-7 How much work is required of an agent to raise a body of mass m from the ground to a height h above the ground? Assume that the body does not accelerate.

Solution In order to raise the body, he must exert an upward force mg. Thus, according to (7-16), the work required is

$$W = \int_0^h mg\,dx = mg\int_0^h dx = mgx\Big|_0^h = mgh$$

where the second inequality follows since the required force mg is a constant. Thus, as was previously shown, in order to raise a 5-kg mass a distance of 5 meters requires the work

$$W = mgh = 5\ \mathrm{kg} \times 9.8\ \frac{\mathrm{m}}{\mathrm{s}^2} \times 5\ \mathrm{m} = 245\ \mathrm{J}$$

Example 7-8 How much work is carried out in stretching a spring of constant k from its equilibrium position by an amount a?

Solution Let us select the coordinate system shown in Figure 7-12, with the origin at the original equilibrium position. When the spring has been stretched a distance y, the spring pulls on the agent with a force $-ky$, and thus he must be pulling down on it with the force $+ky$. Hence the force $F(y)$ is

$$F(y) = ky$$

and therefore the work W required to stretch the spring from $y=0$ to $y=a$ is

$$W=\int_0^a ky\,dy = k\int_0^a y\,dy = \frac{ky^2}{2}\bigg|_0^a$$
$$=\frac{ka^2}{2}$$

where the third equality follows since $dy^2/dy = 2y$.

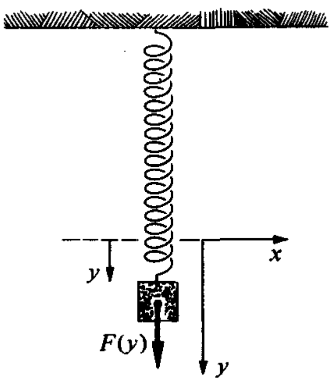


Figure 7-12

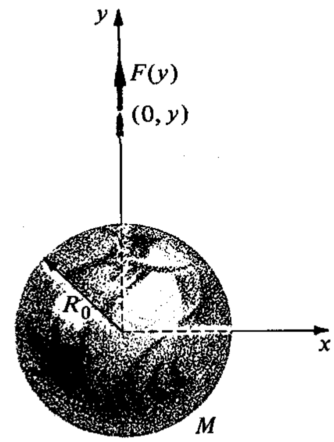


Figure 7-13

Example 7-9 Calculate the minimum work required to raise a rocket of mass m from the surface of a planet of radius R_0 and of mass M to a point a distance R ($>R_0$) from its center. Assume that the rocket moves radially outward at all times.

Solution Let us set up a coordinate system, as in Figure 7-13, with the origin at the center of the planet, and suppose that the direction of motion of the rocket is along the positive y-axis. The force that will produce the minimal work is the one that just compensates for the gravitational attraction due to the planet. According to the law of gravitation, then, at an instant when the rocket is at a distance y from the center of the planet, the agent must exert the force

$$\mathbf{F} = \mathbf{j}\frac{GMm}{y^2}$$

where $\mathbf{j}$ is a unit vector along the y-axis and G is the gravitational constant. Thus the work W required to take the rocket from the point $y=R_0$ to $y=R$ ($>R_0$) is

$$W=\int_{R_0}^{R} F(y)\,dy = \int_{R_0}^{R}\frac{GMm}{y^2}\,dy = GMm\int_{R_0}^{R}\frac{dy}{y^2} = GMm\left(-\frac{1}{y}\right)\bigg|_{R_0}^{R}$$
$$= GMm\left(\frac{1}{R_0}-\frac{1}{R}\right)$$

where the fourth equality follows since $d(1/y)/dy = -1/y^2$.

The work W_∞ required to take the rocket from the surface of the planet to a point very far from it is

$$W_\infty = \frac{GMm}{R_0}$$

and this is obtained by assuming $R \gg R_0$.

7-6 Geometric considerations

In this section we generalize the above considerations by lifting the restriction that the particle be confined to motion along a straight line.

Consider, in Figure 7-14, a particle moving along an arbitrary trajectory AB under the action of various forces, not all of which have been explicitly displayed. The work ΔW carried out by one of these forces $\mathbf{F}$ when the particle undergoes an infinitesimal displacement $\Delta\mathbf{r}$ is, according to (7-7), given by

$$\Delta W = \mathbf{F} \cdot \Delta\mathbf{r} \tag{7-17}$$

Or, equivalently, in terms of the angle θ between the direction of the force $\mathbf{F}$ and the displacement $\Delta\mathbf{r}$, this is $F|\Delta\mathbf{r}| \cos\theta$.

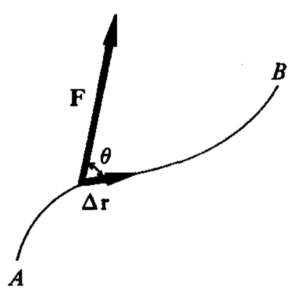


Figure 7-14

As in the corresponding one-dimensional case, in order to calculate the total work $\overline{W}_{AB}$ carried out by the force $\mathbf{F}$ as the particle goes from A to B, imagine the path in the figure broken up into a sequence of infinitesimal displacements $\{\Delta\mathbf{r}\}$. The work $\overline{W}_{AB}$ is then

$$\overline{W}_{AB} = \Sigma\mathbf{F} \cdot \Delta\mathbf{r} \tag{7-18}$$

where the summation is to be carried out over each elementary displacement $\Delta\mathbf{r}$ into which the path has been divided. As in the corresponding one-dimensional considerations, $\overline{W}_{AB}$ depends, in general, on the magnitudes of the elementary displacements $\{\Delta\mathbf{r}\}$. And for the same reason as before, it is convenient to define a new quantity, W, to be the limit of the sum, in (7-18), as the lengths of the infinitesimal displacements $\{\Delta\mathbf{r}\}$ tend to zero. Thus

$$W = \lim_{|\Delta\mathbf{r}|\to 0} \overline{W}_{AB} \tag{7-19}$$

and, from now on, the term "work" will be reserved for this quantity W. Evidently, (7-18) and (7-19) are equivalent to (7-12) and (7-13) for the special case that the path connecting A and B is a straight line.

As for the corresponding one-dimensional case, the limit of the sum in (7-18) and (7-19) defines what is known as the *line integral* of the force $\mathbf{F}$ along the path AB in Figure 7-14. By definition, then,

$$W = \int_A^B \mathbf{F} \cdot d\mathbf{r} \tag{7-20}$$

where the quantity on the right-hand side is called the *line integral* between the points A and B and is defined by the limit of the sum in (7-19). It is easy to confirm for the special case for which the path is the straight line joining the endpoints A and B that the line integral in (7-20) reduces to that of the ordinary definite integral discussed in Section 7-5. Note that, in general, a line integral depends not only on the endpoints A and B but on the particular path connecting these points as well. That is, the values associated with various line integrals along different paths but connecting the same two endpoints are, in general, different.

7-7 The work-energy theorem

Consider, in Figure 7-15, a particle of mass m moving along a trajectory from P_1 to P_2 under the action of a certain force $\mathbf{F}$. By contrast to the preceding sections, in which $\mathbf{F}$ represented a particular one of the forces acting on the body, suppose now that $\mathbf{F}$ represents the *total* force acting on the body. This means that $\mathbf{F}$ is related to the acceleration of the body in accordance with Newton's second law

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} \tag{7-21}$$

There is a very important relation between the work carried out by the total force $\mathbf{F}$ as the particle goes from P_1 to P_2, and its velocity at these two

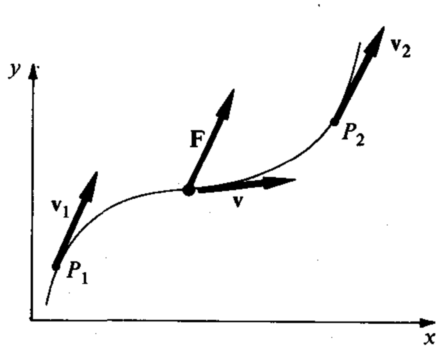


Figure 7-15

points. This relationship is known as the *work-energy theorem* and can be stated as follows:

The work W carried out by the total force **F** *acting on a body as it goes from a point P_1 to the point P_2 is*

$$W = \frac{1}{2} m\mathbf{v}_2^2 - \frac{1}{2} m\mathbf{v}_1^2 \tag{7-22}$$

where $\mathbf{v}_2$ *and* $\mathbf{v}_1$ *are the velocities of the particle at the points P_2 and P_1, respectively and where* $\mathbf{v}^2 \equiv \mathbf{v} \cdot \mathbf{v} = v^2$.

This theorem, whose validity will be established below, not only is of interest in its own right, but also will play a crucial role in the next chapter in connection with our studies of energy conservation.

The quantity $m\mathbf{v}^2/2$ which appears in (7-22) and will play an increasingly important role from now on, has been given the special name *kinetic energy.* The word *kinetic* here means "of or pertaining to motion," and so kinetic energy means the energy associated with motion. At an instant when a particle of mass m has a velocity **v**, its kinetic energy K is thus defined by

$$K = \frac{1}{2} m\mathbf{v}^2 \tag{7-23}$$

Let us note here parenthetically that (7-23) is valid only for velocities that are small compared to the speed of light c. In Section 7-9 the case of $v \sim c$ will be briefly discussed, and the appropriate generalization of this formula will be presented.

The units of kinetic energy are the same as those of work, so the joule is also the unit of kinetic energy. If, for example, a man of mass 75 kg runs at a speed of 2 m/s his kinetic energy is

$$\begin{aligned} K &= \frac{1}{2} mv^2 = \frac{1}{2} \times 75 \text{ kg} \times (2 \text{ m/s})^2 \\ &= 150 \text{ J} \end{aligned}$$

Similarly, the kinetic energy of a proton of mass $m = 1.7 \times 10^{-27}$ kg traveling at a speed of 10^6 m/s is

$$\begin{aligned} \frac{1}{2} mv^2 &= \frac{1}{2} \times (1.7 \times 10^{-27} \text{ kg}) \times (10^6 \text{ m/s})^2 \\ &= 8.5 \times 10^{-16} \text{ J} \end{aligned}$$

To establish the validity of the work-energy theorem in (7-22), consider first the special case of a particle confined to motion along the x-axis of a

coordinate system and under the action of the force $F = F(x)$. Multiplying both sides of the equation of motion

$$F(x) = m\frac{dv}{dt}$$

by $v\,dt$, we obtain

$$F(x)\,dx = mv\,dv$$

where the left-hand side follows since $dx = v\,dt$ and the right-hand side follows since $(v\,dt)dv/dt = v\,dv$. Integrating both sides from the point x_1, where the particle has the velocity v_1, to the point x_2, where its velocity is v_2, we obtain (since m is a constant)

$$\int_{x_1}^{x_2} F(x)\,dx = m\int_{v_1}^{v_2} v\,dv$$

Now, according to (7-16), the left-hand side here represents the work W carried out on the particle by $F(x)$ as it goes from x_1 to x_2. Hence, using the fact that $v\,dv = d(\frac{1}{2}v^2)$, we may integrate the right-hand side and thus write

$$\begin{aligned} W = m\int_{v_1}^{v_2} v\,dv &= \frac{1}{2}mv^2\Big|_{v_1}^{v_2} \\ &= \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2 \end{aligned}$$

This then establishes the validity of this theorem in the one-dimensional case.

The extension of this argument to the three-dimensional case is straightforward. This time we take the scalar product of both sides of Newton's law in (7-21) with the quantity $\mathbf{v}\,dt$. In terms of the displacement $d\mathbf{r}$ of the particle in a time interval dt, we have $d\mathbf{r} = \mathbf{v}\,dt$ and thus

$$\begin{aligned} \mathbf{F}\cdot d\mathbf{r} &= m\frac{d\mathbf{v}}{dt}\cdot\mathbf{v}\,dt \\ &= m\mathbf{v}\cdot d\mathbf{v} \end{aligned} \tag{7-24}$$

where the second equality follows since $(\mathbf{v}\,dt)\cdot d\mathbf{v}/dt = \mathbf{v}\cdot d\mathbf{v}$. Therefore, using the fact that $d(\mathbf{v}^2) = d(\mathbf{v}\cdot\mathbf{v}) = 2\mathbf{v}\cdot d\mathbf{v}$ we find, on integrating (7-24), that

$$\begin{aligned} \int_{P_1}^{P_2} \mathbf{F}\cdot d\mathbf{r} &= \int_{\mathbf{v}_1}^{\mathbf{v}_2} \frac{1}{2}md(\mathbf{v}^2) \\ &= \frac{1}{2}m(\mathbf{v}_2^2 - \mathbf{v}_1^2) \end{aligned}$$

Finally, since the integral on the left represents the work W carried out by $\mathbf{F}$, the validity of (7-22) for the more general three-dimensional case is established.

The work-energy theorem can also be used to define what is meant by the work carried out by a force $\mathbf{F}$ that depends explicitly on time t or on the

velocity **v** of the particle. The work W carried out by such a force in taking a particle from a point where it has a velocity $\mathbf{v}_1$ to one where it has a velocity $\mathbf{v}_2$ is *defined* to be the difference in the kinetic energies of the particle at these two points. That is, for this more general class of forces, W is defined so that the work-energy theorem is satisfied. Thus, although the concept of work has been defined only for a constant or a spatially varying force, from now on we shall use the concept of work in this more general sense. In other words, the work-energy theorem is not restricted to forces of the form $F(x)$ or $\mathbf{F}(\mathbf{r})$, but is very generally valid for all types of forces.

Example 7-10 A body of mass 2 kg moves under the action of a force **F**. If initially it has a velocity $v_1 = 1$ m/s and 5 seconds later its velocity is 1.5 m/s, calculate the kinetic energy of the body at these two instants and the amount of work carried out by **F**.

Solution Since at the initial instant the velocity of the body is 1 m/s, it follows from (7-23) that its kinetic energy K_0 then is

$$K_0 = \frac{1}{2}mv_1^2 = \frac{1}{2} \times 2\text{ kg} \times (1\text{ m/s})^2$$
$$= 1\text{ J}$$

Similarly, the kinetic energy K_5, 5 seconds later, is

$$K_5 = \frac{1}{2}mv_2^2 = \frac{1}{2} \times 2\text{ kg} \times (1.5\text{ m/s})^2$$
$$= 2.25\text{ J}$$

According to the work-energy theorem, the work W carried out by **F** during this time interval is

$$W = K_5 - K_0 = 2.25\text{ J} - 1\text{ J}$$
$$= 1.25\text{ J}$$

Example 7-11 A body of mass m is thrown vertically upward with an initial speed v_0. What is its velocity when it has risen a distance y?

Solution While the body is in flight, the only force acting on it is the force of gravity, which exerts a force mg directed vertically downward. Accordingly, the total work carried out on the body when it has risen a distance y is $-mgy$. If v is the velocity of the body at the instant when it is at the elevation y, then, according to the work-energy theorem, (7-22),

$$-mgy = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$$

Canceling the common factor m and multiplying throughout by 2, we obtain

$$v^2 = v_0^2 - 2gy$$

and this agrees in all important respects with the corresponding formula in (2-18), as it must.

Example 7-12 If a particle of mass m vibrating at the end of a spring of constant k has a velocity v_0 when it goes through the equilibrium point, calculate the amplitude of its motion A in terms of v_0.

Solution As the particle travels from the equilibrium position to the point of maximum amplitude A, where its velocity vanishes, the spring exerts a force on it that is always directed toward the initial equilibrium position. According to the result of Example 7-8, the work carried out *on* the spring in this process is $kA^2/2$. It thus follows that the work carried out by the spring on the particle as it goes from the equilibrium position to the point corresponding to the maximum displacement, A, is $-kA^2/2$. Furthermore, at the position of maximum displacement the velocity of the particle vanishes, and therefore, according to the work-energy theorem,

$$-\frac{1}{2}kA^2 = 0 - \frac{1}{2}mv_0^2$$

Solving for A we find the sought-for relation between the amplitude A and the velocity v_0

$$A = \sqrt{\frac{m}{k}}$$

and this agrees with (6-16), as it must.

7-8 Power

If an agent exerts a force on a body so that it moves from one place to another, then in addition to the work carried out in this process, very often we are also interested in the time it takes for this work to be accomplished. An agent who carries out 2 joules of work in 1 second is radically different from one who requires a full year for this task. In order to give a quantitative measure to this concept which involves both work and the time required to achieve it, it is convenient to introduce the notion of *power.*

Power is the *time rate of doing work.* Suppose that as a consequence of the action of a certain force $\mathbf{F}$ a body undergoes a small displacement $\Delta\mathbf{r}$ during a time interval Δt. According to (7-17), the work ΔW carried out is $\mathbf{F} \cdot \Delta\mathbf{r}$ and the *average* power $\bar{P}$ expended by $\mathbf{F}$ during this time interval is defined by

$$\bar{P} = \frac{\Delta W}{\Delta t} \tag{7-25}$$

In physical terms then, $\bar{P}$ is a measure of the work carried out per unit time.

A quantity related to $\bar{P}$ is the *instantaneous power,* or the *power* P expended by the force $\mathbf{F}$. This is defined to be the limit of the average power $\bar{P}$ as the length of the time interval Δt tends to zero. It follows, by use of (7-25) and the definition of a derivative, that P may be expressed by

$$P = \frac{dW}{dt} \tag{7-26}$$

Or, equivalently, since $\Delta W = \mathbf{F} \cdot \Delta \mathbf{r}$, we find by substituting this formula into (7-25) and taking the limit $\Delta t \to 0$ that

$$P = \mathbf{F} \cdot \mathbf{v} \tag{7-27}$$

where $\mathbf{v} = d\mathbf{r}/dt$ is the instantaneous velocity of the particle. Thus power can be measured by the product of the instantaneous velocity of the particle and the component along the direction of motion of the force acting on it.

According to the definition in (7-26), the unit of power is that of work per unit time. Hence the unit of power is the joule per second (J/s). Other units of power are the erg per second, the foot-pound per second, and the "horsepower," which is defined to be 550 ft-lb/s. In SI units there is also the unit of power called the *watt* (W), which is defined to be 1 joule per second:

$$1 \text{ W} = 1 \text{ J/s} \tag{7-28}$$

For example, a 100-watt bulb is one that consumes 100 joules during each second that it is turned on.

Example 7-13 Show that the power P expended on a particle by the total force $\mathbf{F}$ acting on it is related to its kinetic energy K by

$$P = \frac{dK}{dt} \tag{7-29}$$

Solution Since $\mathbf{F}$ is the total force acting on the particle it follows, by taking the dot product of $\mathbf{v}$ with Newton's second law, that

$$\mathbf{F} \cdot \mathbf{v} = m\mathbf{v} \cdot \frac{d\mathbf{v}}{dt}$$

But the right-hand side is the time derivative of the kinetic energy $mv^2/2$. Therefore, since the left-hand side is the power P according to (7-27), the desired result in (7-29) follows.

It is of interest to note that, because of (7-26), this result in (7-29) may be expressed as

$$\frac{dW}{dt} = \frac{dK}{dt} \tag{7-30}$$

In other words, the rate at which the total force acting on a body carries out work on it is the same as the rate at which its kinetic energy increases. This is evidently another way of stating the work-energy theorem in (7-22).

Example 7-14 A particle of mass m falls from a height h above the ground toward the earth.

(a) At a time t after the particle has started to fall to the earth, what work has the earth's gravitational field carried out on the particle?

(b) What power is being expended at this instant?

(c) Assuming $m = 1$ kg and $h = 25$ meters, find numerical values for these two quantities at the instant the particle hits the ground.

Solution

(a) In a coordinate system with y-axis vertically upward and with origin on the ground, the position and velocity at any time t is, according to (2-16) and (2-17),

$$y(t) = h - \frac{1}{2} g t^2 \tag{7-31}$$

$$v(t) = -gt \tag{7-32}$$

Now when the particle is at a distance y above the ground, the work that has been carried out on it by the gravitational field of the earth is

$$W = \int_h^y (-mg)\, dy = -mgy \Big|_h^y$$
$$= mg\,(h - y)$$

Making use of (7-31), this formula for W may be reexpressed as

$$W = mg\,(h - y) = \frac{1}{2} mg^2 t^2$$

(b) According to (7-26), the rate of doing work is obtained by differentiating W with respect to time. The result is

$$P = \frac{dW}{dt} = \frac{d}{dt}\left(\frac{1}{2} mg^2 t^2\right)$$
$$= mg^2 t$$

Or, equivalently, P may be obtained by use of (7-27). Since $F = -mg$ and $v = -gt$, the same result as above follows.

(c) The particle hits the ground at the instant t_0 when $y = 0$ or at time

$$t_0 = \sqrt{\frac{2h}{g}}$$

Substituting this value into the above formulas for W and P we find that

$$W = \frac{1}{2} mg^2 t_0^2 = \frac{1}{2} mg^2 \frac{2h}{g} = mgh$$

and

$$P = mg^2 t_0 = mg^2 \sqrt{\frac{2h}{g}}$$

which, for the values $h = 25$ meters and $m = 1$ kg, are

$$W = 245\ \text{J}$$

$$P = 217\ \text{W}$$

†7-9 Relativistic kinetic energy

Since Newton's laws of motion are not valid at speeds comparable to that of light ($\sim 3.0 \times 10^8$ m/s), we might expect that the definition of kinetic energy in

(7-23) is valid only at low speeds. This is indeed the case. The purpose of this section is to describe briefly the correct high-velocity form for this important dynamical quantity.

Although Newton's second law is not correct at high speeds, studies in relativistic mechanics show that there is nevertheless a form of the work-energy theorem that is applicable at all observable speeds less than the speed of light c. The form of this theorem is precisely the same as that in (7-22) or (7-30), provided that the kinetic energy K in these formulas is replaced by the relativistic energy ϵ. For the case of a particle traveling at a velocity v, the relativistic energy ϵ is

$$\epsilon = \frac{mc^2}{(1 - v^2/c^2)^{1/2}} \tag{7-33}$$

where m is the mass of the particle. The significance of this relativistic quantity ϵ as defined here is that if it is substituted into either (7-22) or (7-30), the correct form of the relativistic work-energy theorem is obtained. The fact that (7-33) predicts that $\epsilon \to \infty$ as $v \to c$ and that ϵ is imaginary for $v > c$ is no difficulty, since experiment shows that no particles ever travel at speeds greater than c.

Figure 7-16 shows a plot of ϵ in (7-33) as a function of the parameter v/c. According to the graph, for very small values of v/c, ϵ has the constant value mc^2 and as v approaches the speed of light c, it becomes infinitely large. To achieve a value of $2mc^2$ for ϵ, for example, it is necessary, according to (7-33), that the particle travel at a speed $\cong 0.87c$.

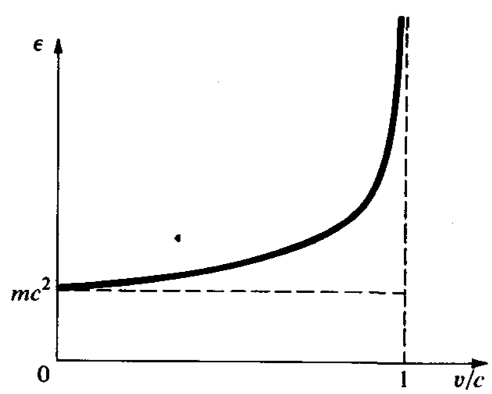


Figure 7-16

In order to relate the relativistic energy ϵ in (7-33) to the nonrelativistic form of kinetic energy in (7-23), let us make use of the binomial expansion

$$\frac{1}{(1 - x^2)^{1/2}} \cong 1 + \frac{1}{2}x^2 + \cdots$$

which is valid for x sufficiently small. Making use of this result, we may

approximate (7-33) for small v/c by

$$\begin{aligned}\epsilon &= mc^2\frac{1}{(1-v^2/c^2)^{1/2}}\\ &\cong mc^2\left(1+\frac{1}{2}\frac{v^2}{c^2}+\cdots\right)\\ &= mc^2+\frac{1}{2}mv^2+\cdots\end{aligned} \tag{7-34}$$

Hence up to the additive constant mc^2, in the low velocity limit, the form for ϵ in (7–33) agrees with the definition for kinetic energy in (7-23). The term mc^2 is known as the *rest energy* of the particle and is the basis of the relation

$$E = mc^2$$

which relates the energy content E of matter with its mass m. We shall return to this relation between energy and mass in more detail in Chapter 9.

Reference to (7-34) shows that, for small speeds, the velocity dependence of the relativistic energy ϵ in (7-33) is the same as is the nonrelativistic kinetic energy in (7-23). This suggests that the relativistic *kinetic energy* be defined as the difference between the relativistic energy ϵ and the rest energy mc^2; that is,

$$K = mc^2\left\{\frac{1}{(1-v^2/c^2)^{1/2}}-1\right\} \tag{7-35}$$

Reference to (7-34) shows that this form for the kinetic energy agrees with the corresponding nonrelativistic form (7-23) at low speeds. Hence (7-35) is the correct form for the kinetic energy of a particle and is applicable at all speeds.

7-10 Summary of important formulas

Given two vectors **A** and **B**, their *dot product* is defined by

$$\mathbf{A}\cdot\mathbf{B} = AB\cos\theta \tag{7-4}$$

where θ is the angle between **A** and **B**. In terms of this notation, if a body undergoes an infinitesimal displacement $\Delta\mathbf{r}$ under the action of a force **F**, the work ΔW carried out is

$$\Delta W = \mathbf{F}\cdot\Delta\mathbf{r} \tag{7-17}$$

The work W carried out by a force **F** (depending only on position) along the path AB in Figure 7-14 is given by a sum of terms in the form of (7-17), one for each path element $\Delta\mathbf{r}$ into which AB is subdivided. In the limit $|\Delta\mathbf{r}|\to 0$, W becomes the line integral

$$W = \int_A^B \mathbf{F}\cdot d\mathbf{r} \tag{7-20}$$

and for the special case in which the path is a straight line, this is the definite integral

$$W = \int_{x_1}^{x_2} F(x)\,dx \tag{7-16}$$

where x_1 and x_2 are the endpoints.

The basic result of this chapter is the work-energy theorem:

$$W = \frac{1}{2}m\mathbf{v}_2^2 - \frac{1}{2}m\mathbf{v}_1^2 \tag{7-22}$$

where W is the work carried out on a body by the total force acting on it as it goes from the point P_1 to the point P_2, and $\mathbf{v}_1$ and $\mathbf{v}_2$ are the respective velocities at these two points. The kinetic energy of a body of mass m when traveling at a velocity $\mathbf{v}$ is

$$K = \frac{1}{2}m\mathbf{v}^2 \tag{7-23}$$

QUESTIONS

1. Define briefly what is meant by the following terms: (a) work; (b) kinetic energy; (c) constant of the motion; (d) work-energy theorem; and (e) power.
2. Explain why it is that if you raise a body vertically through a certain height, then the work you do will *differ* depending on how you raise it.
3. In light of Question 2, is there a minimum amount of work that you can do in raising a body through a given height? Is there a maximum amount of work you can do in raising a body? Explain.
4. For what class of forces have we *defined* the concept of work? Does this definition of work apply to a viscous force proportional to the velocity v of a body? Does it apply to frictional forces?
5. How is the work carried out by a velocity-dependent force defined? Does the work-energy theorem apply to the work carried out by such a force?
6. Consider the bob of a pendulum swinging back and forth. Does the tension in the string carry out any work on the bob? Explain.
7. A small body slides down a *smooth*, hemispherical bowl whose open end faces upward. What are the forces acting on the body? Which of these forces carry out work on it while it is in motion?
8. Assuming that the earth orbits about the sun in a circular orbit, does the gravitational force of the sun carry out any work on the earth? Does the kinetic energy of the earth change during the course of a year under these circumstances? Explain.
9. Since the earth actually goes about the sun in an elliptical path, explain why the gravitational force of the sun carries out work on it. Does this mean that the kinetic energy of the earth varies during the year?
10. In Figure 7-3 explain why the forces on the string due to the pegs at C_1 and C_2 carry out no work on the string. Is the gravitational force on the bob the only force carrying out work on it? Explain.
11. Suppose you pick up a bag of sand

from the ground, raise it through a certain height, and place it on a shelf. What is the work carried out on it according to the work-energy theorem? How do you account for the fact that you actually had to carry out work in this process?

12. A block slides along a rough, horizontal surface and eventually comes to rest. What is the sign of the work carried out on the block according to the work-energy theorem? What can you say about the direction of action of the force acting on the block based on this information?
13. If the force acting on a moving object has a positive component along the direction of motion, would you expect the kinetic energy of the block to increase or to decrease? Is the work carried out by the force positive or negative in this case? Check your answer with the prediction of the work-energy theorem.
14. A rubber ball dropped from a certain height h bounces on a horizontal floor, and rises to a height $h/2$. What has happened to the kinetic energy of the ball as a result of the bounce? What is the sign of the work carried out on the ball during the bounce? What force has carried out this work?
15. A person standing still suddenly starts to run and acquires a certain speed v_0. What agent has carried out the work to produce this change in kinetic energy? What is the direction of the force relative to the direction of motion and what is the sign of the work?

PROBLEMS

1. A block is pulled along a rough horizontal road by a horizontal force of 3 newtons. If it travels at a uniform speed of 2 m/s for a distance of 10 meters, how much work is carried out by this force on the block? How much work is carried out by the frictional force that also acts on the block?
2. A monkey of mass 15 kg climbs a vertical rope at a constant speed of 3 m/s. How much work has he carried out in a 2-second interval?
3. If the monkey in Problem 2 climbs upward at an acceleration of 1 m/s^2 (a) what upward force must the rope exert on the monkey and (b) how much work does he carry out if he climbs a distance of 5 meters?
4. A 500-gram block at rest on a rough, horizontal surface is suddenly given a push so that it starts to travel at a speed of 2 m/s. It comes to rest in a distance of 3 meters.
   (a) What was its (presumed constant) acceleration?
   (b) What was the strength of the frictional force acting on it?
   (c) What work was carried out on the block by this frictional force?
5. An elevator of mass 2.0×10^4 kg descends with a uniform acceleration of 1.5 m/s^2.
   (a) What is the tension in the supporting cable?
   (b) How much work does the supporting cable carry out on the elevator while it descends a distance of 15 meters?
   (c) How much work is carried out by gravity during this 15 meter descent?
6. Consider again the elevator of Problem 5, but suppose that initially it is rising upward at a speed of 6 m/s and has a constant downward acceleration of 1 m/s^2.

(a) How far does the elevator travel before it comes to rest?
(b) How much work does the supporting cable carry out on the elevator during the time that it travels in (a)?
(c) Compare the result of (b) with the work carried out by the gravitational force of the earth during the same time interval.

7. A block slides down a smooth, inclined plane of elevation angle α. Show that the work carried out by the gravitational force of the earth on the block when it travels a distance l down the plane is the same as if the block had dropped through a vertical height $l \sin \alpha$.

8. Consider the two vectors **A** and **B**:

$$\mathbf{A} = 3\mathbf{i} - 2\mathbf{j} \qquad \mathbf{B} = 2\mathbf{i} + \mathbf{j}$$

where **i** and **j** are unit vectors along the axes of a given coordinate system.
(a) What are the magnitudes of **A** and **B**?
(b) Find a unit vector parallel to **B**.
(c) Calculate the scalar product of **A** and **B**.

9. For the two vectors **A** and **B** in Problem 8, calculate (a) the component of **A** along the direction of **B** and (b) the component of **B** along the direction of **A**.

10. Find the angle between the two vectors $(4\mathbf{i}+3\mathbf{j})$ and $(\mathbf{i}-\mathbf{j})$ by calculating their scalar product and their magnitudes.

11. Find the value of the parameter α so that the vector $2\mathbf{i}+\alpha\mathbf{j}$ is perpendicular to the vector $3\mathbf{i}+2\mathbf{j}$. For what value of α are these two vectors parallel? (*Hint*: Two vectors are parallel if their scalar product is the same as the product of their magnitudes.)

12. A boy pushes a lawnmower across a lawn. If, as shown in Figure 7-17, he exerts a force of 100 newtons directed so that it makes an angle of 45° with the horizontal, how much work does he carry out if he pushes the mower a distance of 20 meters?

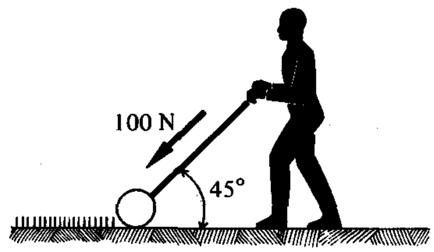


Figure 7-17

13. If the force $\mathbf{F}_0$ in Figure 7-4 has a strength of 50 newtons and makes an angle of 30° with the horizontal, how much work is carried out if the block is displaced horizontally a distance of 10 meters?

14. A particle moving along the x-axis of a certain coordinate system is subject to a force $F(x)$, given by

$$F = 2x^2 - x^3$$

where F is in newtons and x is in meters.
(a) How much work is carried out by this force if it moves the particle a distance of 2 meters in the direction of the positive x-axis starting from the origin?
(b) How much work is carried out by this force if the particle is moved from the point $x =$ 3 meters to the point $x =$ 1 meter?

15. For the situation described in Problem 14, find the distance from the origin that the particle must be moved along the x-axis so that the force $F(x)$ does no work.

16. A particle moving along the x-axis of a certain coordinate system is subject to the force

$$F(x) = \alpha x + \beta$$

where

$$\alpha = 1\ \text{N/m} \quad \text{and} \quad \beta = 2\ \text{N}$$

(a) Calculate the work carried out by this force as the particle is displaced from the origin to the point $x = 2$ meters.
(b) Make a plot of $F(x)$ as a function of x and compute geometrically the area under this curve between the vertical lines erected at $x = 0$ and at $x = 2$ meters. Compare with your result from (a) and account for any differences.

17. What is the minimum amount of work that must be carried out to raise a rocketship with a mass of 4.0×10^4 kg from the surface of the earth to a height of 500 km? Assume that the earth is a sphere of radius 6.4×10^3 km and that the rocket moves radially outward at all times.

18. Calculate the kinetic energies in joules and in electron volts of the following:
(a) An electron of mass 9.1×10^{-31} kg in a hydrogen atom where it has a speed of 2.19×10^6 m/s.
(b) An automobile of mass 2000 kg traveling at 90 km/hr.
(c) The earth ($M = 6.0 \times 10^{24}$ kg) as it travels about the sun at a speed of 30 km/s.

19. A particle of mass 0.1 kg moves along the trajectory $x(t) = 3t + t^3$, where x is in meters and t is in seconds.
(a) What is the kinetic energy of the particle at time t?
(b) What is the acceleration of the particle and the strength of the force acting on it at this instant?
(c) How much work does this force carry out on the particle in the 2-second interval from $t = 1$ second to $t = 3$ seconds?

20. A 5-kg object travels at a certain speed so that its kinetic energy is 20 joules. **(a)** What is the speed? **(b)** If it is now subjected to the action of a certain force so that its speed is doubled, how much work does this force carry out?

21. A 15-gram bullet is traveling at a speed of 1.5 km/s when it enters a bag of sand and comes to rest in a distance of 10 cm. **(a)** How much work does the sand carry out on the bullet? **(b)** Assuming that the force producing this work is constant, calculate the strength of the force.

22. A 100-gram ball is traveling at a speed of 25 m/s when it strikes a wall. If its velocity upon bouncing from the wall is 15 m/s **(a)** what was its initial kinetic energy and **(b)** how much work was carried out on the ball during its collision with the wall?

23. A block of mass m is projected with an initial speed v_0 down an inclined plane of angle α and of length l. Assume that the plane is rough and that the block comes to rest just as it reaches the bottom of the plane.
(a) How much work has the gravitational field of the earth carried out on the block during its motion?
(b) What is the total work W carried out on the block during its motion?
(c) Based on your results to (a) and (b), calculate the work carried out on the block by the force of friction.

24. A particle of mass m is attached to a spring of constant k, the entire system being immersed in a viscous fluid. Suppose that you displace the particle from its equilibrium position by a distance a and find on releasing it at rest that when it returns to its equilibrium position its velocity vanishes.
(a) What was the total work carried out on the particle during its

journey from maximum displacement to equilibrium?

(b) How much work was carried out by the spring during this process?

(c) From your results for (a) and (b), calculate the work carried out by the viscous force on the particle.

25. An asteroid of mass 10^7 kg is at a distance of 10^5 km from the center of the earth. Assume that it is at rest initially at this point and then falls to the earth. **(a)** Make use of the results of Example 7-9 to calculate the work that the force of gravity carries out on it as it descends toward the surface of the earth. **(b)** By use of the work-energy theorem, calculate its velocity when it reaches the earth's surface.

26. Show, by use of the results of Example 7-9 and the work-energy theorem, that the minimum velocity v_m with which a body must be ejected from the surface of a planet of mass M so that it escapes its influence is

$$v_m = \left(\frac{2GM}{R_0}\right)^{1/2}$$

where R_0 is the radius of the planet. (*Note*: The velocity v_m as here defined is called the *escape velocity*. See Problem 22 in Chapter 5.)

27. A bead of mass m travels around a rough, horizontal, circular hoop of radius R. Suppose its initial velocity is v_0, and after the completion of one turn, this has been reduced to $\frac{1}{2}v_0$. **(a)** Calculate the work carried out on the bead by the force of friction. **(b)** Assuming that the frictional force is constant, calculate its strength.

28. Consider a particle of mass m confined to motion along a single direction. Starting with (7-23), show that the time derivative of the kinetic energy of the particle is the same as the rate at which the forces acting on the particle do work on it.

29. Calculate the power output of the monkey in Problem 2.

30. For the situation described in Problem 1, at what *rate* does the force carry out work on the block?

31. A 2-kg block has an initial velocity of 2 m/s and is subject to the action of a certain force so that it undergoes an acceleration of $-10\ \mathrm{m/s^2}$.

(a) How far does it travel before coming to rest?

(b) What is the strength of the force acting on it?

(c) Calculate the power being expended on the block by the force at any time t.

32. A 2-kg particle moves along the trajectory $x = 2t^2 + t^4$, where x is in meters and t is in seconds.

(a) What is the acceleration at time t?

(b) What is the force on the particle at time t?

(c) Calculate the rate at which work is being carried out on the particle at any time t.

(d) How much work is carried out on the particle by the force field in the time interval from 0 to 3 seconds?

33. Consider a particle of mass m moving in one dimension under the action of a force field $F(x)$. Let $V(x)$ be a function defined so that

$$F(x) = -\frac{dV}{dx}$$

Making use of the work-energy theorem, show that the quantity

$$\frac{1}{2}Mv^2 + V(x)$$

is a constant of the motion; that is, show that this quantity does not change in the course of time.

34. Consider, in Figure 7-18, a particle of mass m sliding down a smooth

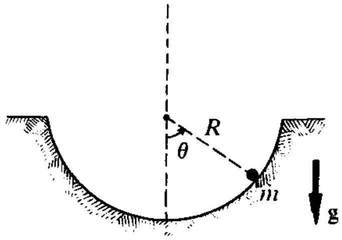


Figure 7-18

hemispherical hole of radius R. **(a)** Show that the work dW carried out by all of the forces acting on the particle when it undergoes a small angular displacement $d\theta$, where θ is defined in the figure, is

$$dW = -mg \sin\theta R\, d\theta$$

(b) Show therefore that the total work W carried out on the particle as it goes from the top of the hole to its bottom is Rmg. **(c)** Making use of the result in (b), calculate the velocity of the particle at the bottom of the hole assuming that it is released from rest at the top.

35. Let W be the work required for a certain force $F(x)$ to move a body from some reference point x_0 to the point x; that is,

$$W = \int_{x_0}^{x} F(x')\, dx'$$

Show by applying the rules for differentiation directly to this formula that the power dW/dt produced by this force is $F(x)v$.

†36. Consider the one-dimensional motion of a particle in a force field $F(x)$ from the viewpoint of an observer who has the constant acceleration a_0 relative to a Newtonian frame. Starting with the equation of motion

$$m\frac{dv'}{dt} = F - ma_0$$

show that the work-energy theorem in this system is

$$W = \frac{1}{2}m(v_2'^2 - v_1'^2) + ma_0(x_2' - x_1')$$

where x_2' is the final position of the particle when it has the observed velocity v_2', and similarly for x_1' and v_1'.

†37. Generalize the results of Problem 36 to the three-dimensional case. Assume that the relative acceleration $\mathbf{a}_0$ between the systems is still constant.

8 Potential energy and energy conservation

The more things change the more they remain the same.
MONTAIGNE

8-1 Introduction

A *constant of the motion* is a dynamical quantity which, although depending on the positions and velocities of the particles of the system of interest, does not itself change in the course of time. The velocity **v** of an *isolated* particle is an obvious example of a constant of the motion. For according to the first law, if there are no forces acting on the particle it will continue to move with its original velocity so that **v** is a constant in this case. A physical quantity of this type whose value does not change in time is said to be *conserved.* Also, we shall refer to the constancy in time of a physical quantity as the *law of the conservation* of that quantity.

In studies of mechanics, interest is usually centered around three particulary important constants of the motion. Two of these are associated with the laws of the conservation of *linear momentum* and the conservation of *angular momentum*, and these will be discussed in Chapters 9 and 10. In this chapter our main concern will be with that constant of the motion known as the *energy.* In order to simplify matters, we shall first study the associated

law of energy conservation, not in its full generality but rather in the context of a single particle under the influence of a fixed force field. The generalizations necessary to describe composite systems will be discussed in Chapters 11 and 12.

8-2 Potential energy—one dimension

Consider, in Figure 8-1, a body acted upon by a certain force F under whose influence it undergoes one-dimensional motion from a point x_1, where it has a velocity v_1, to the point x_2, where it has the velocity v_2. According to the work-energy theorem, (7-22), the work W carried out by F in this process is the difference

$$W = \frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2 \tag{7-22}$$

where m is the mass of the body. It is important to recall that although (7-22) was derived for force functions that depend only on position (and not on v or t), it is also applicable to other types of forces, for which it serves as the definition of work.

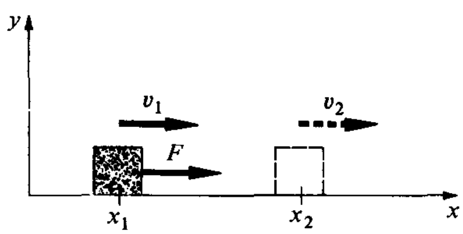


Figure 8-1

A one-dimensional force field F will be said to be *conservative* provided that there exists a function $V(x)$ which depends only on position (and not, for example, on t or v) and is such that if the body undergoes a displacement from x_1 to an arbitrary point x_2 under the action of F, then the work W carried out is

$$W = V(x_1) - V(x_2) \tag{8-1}$$

Note that in order for F to be conservative, (8-1) must be valid for *arbitrary displacements* of the body. The function $V(x)$ is known as the *potential-energy function.* It describes an important attribute of the body when it is subject to the force $F(x)$, known as its *potential energy.* Just as the term *kinetic energy* describes the energy associated with the motion of a body, in the same way *potential energy* characterizes the energy the body has by virtue of its position.

Since the definition of potential energy in (8-1) is in terms of the difference $[V(x_1) - V(x_2)]$ it follows that the function V is itself defined only up to the

addition of an arbitrary constant. Thus, if $V(x)$ is the potential associated with a given force field F, then the function $[V(x)+C]$, where C is an arbitrary constant, is also a potential function associated with the same force field F. In order, therefore, to specify a unique potential energy for a given force F, it is necessary to state how this constant is to be selected. Note, however, that there is no physical significance attached to any one particular choice.

To obtain an explicit relation between a conservative force and its potential energy function, consider a force field $F(x)$ which depends only on position x. Comparison of the definition in (8-1) with (7-14) and (7-15) shows that V and F must be related by

$$F(x) = -\frac{dV}{dx} \tag{8-2}$$

and therefore if $V(x)$ is known, the force function $F(x)$ may be obtained simply by differentiation. Conversely, given $F(x)$, a potential energy function $V(x)$ may be obtained by integrating (8-2); that is,

$$V(x) = -\int_{x_0}^{x} F(x)\,dx \tag{8-3}$$

with x_0 some arbitrary, but fixed, point. (Recall in this connection that the potential function is determined only up to a constant.) Moreover, since this argument is applicable for *all* force functions $F(x)$, it follows from (8-1) and (8-3) that *all one-dimensional* force fields $F(x)$ are necessarily conservative. For it is easy to confirm that any $F(x)$, when substituted into (8-3), will satisfy (8-1).

It is important to note that the arguments leading to (8-2) and (8-3) are applicable only to one-dimensional force fields $F(x)$. As will be illustrated below, *not all* three-dimensional force fields $\mathbf{F}(\mathbf{r})$ that are independent of t and of the velocity $\mathbf{v}$ of the particle are necessarily conservative.

Returning now to the situation of a one-dimensional force field F, it is of interest to ask if it is possible for a force field F which varies explicitly with time t, or with the velocity v of the particle, to be conservative. The answer is that, in general, such force fields are *not* conservative. For although the work W associated with nonconservative forces may under special circumstances be represented by a difference, as in (8-1), this is generally not the case, as illustrated in Example 8-1.

Example 8-1 A particle of mass m is confined to one-dimensional motion under the action of a force $F = \gamma t$ where γ is a positive constant. Assuming that initially the particle is at rest at the point x_0, how much work has this force carried out on the particle at time t when it is located at the point x?

Solution Integrating the equation of motion $\gamma t = m\,dv/dt$ we obtain

$$v(t) = \frac{\gamma t^2}{2m} \tag{8-4}$$

and

$$x(t)=\frac{\gamma t^3}{6m}+x_0 \tag{8-5}$$

since, by assumption, $x(0)=x_0$, and $v(0)=0$. The work W carried out as the particle goes from x_0 to x may be obtained by use of (7-22) to be

$$W=\frac{1}{2}m(v_2^{\,2}-0)=\frac{1}{2}m\left(\frac{\gamma t^2}{2m}\right)^2=\frac{\gamma^2 t^4}{8m}$$

where v_2 is the velocity at time t and where the second equality follows by use of (8-4). Solving (8-5) for t and substituting into this formula for W, we obtain

$$W=\frac{\gamma^2 t^4}{8m}=\frac{\gamma^2}{8m}\left[\frac{6m}{\gamma}(x-x_0)\right]^{4/3}$$

Since this is *not* the difference of a single function evaluated at the points x and x_0, it follows that the force $F=\gamma t$ is not conservative.

Example 8-2 Calculate the force $F(x)$ associated with the following one-dimensional potential energies:

(a) $V=-\alpha x$
(b) $V=\beta x^2-\gamma x^3$
(c) $V=V_0\cos\alpha x$

Assume that the parameters α, β, γ, V_0 are fixed constants in each case.

Solution

(a) Since the derivative of a linear function of x is constant, it follows by substituting this form for V into (8-2) that the force F is the constant α.

(b) Similarly, we find by use of (8-2) that the force $F(x)$ is, in this case,

$$F(x)=-2\beta x+3\gamma x^2$$

(c) Making use of the rules for differentiating the trigometric functions as described in Appendix C, we find by use of (8-2) that

$$F(x)=\alpha V_0\sin\alpha x$$

8-3 Potential energy of simple one-dimensional systems

In this section we shall calculate the potential function V associated with (1) the uniform field, (2) the simple harmonic oscillator, and (3) the inverse-square gravitational field. The results for these three systems are summarized in Table 8-1.

Consider first the case of a particle of mass m confined to one-dimensional motion along the vertical and moving in the uniform gravitational field of the earth. Let us take a coordinate system as in Figure 8-2 so that the force $F(y)$ is $-mg$. According to (8-2), the potential V associated with this force is that function whose negative derivative has the constant value $-mg$. Brief reflection then shows that the potential $V(y)$ for the

Table 8-1 Potential energies of selected force fields

Force field	$F(y)$	$V(y)$
Uniform	$-mg$	mgy
Simple harmonic oscillator	$-ky$	$\frac{ky^2}{2}$
Inverse-square gravitational	$-\frac{GMm}{y^2}$	$-\frac{GMm}{y}$

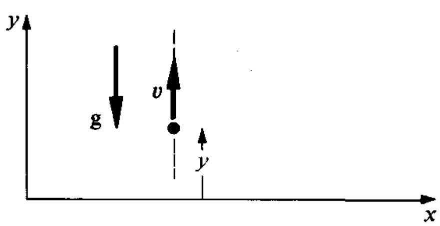


Figure 8-2

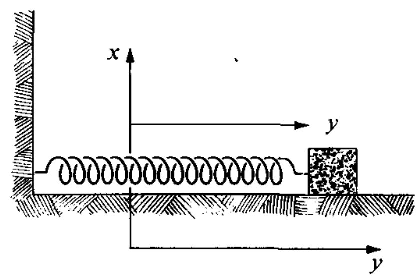


Figure 8-3

uniform gravitational field may be chosen to be

$$V(y) = mgy \tag{8-6}$$

provided that the arbitrary constant is selected so that the potential vanishes at the point $y = 0$.

For the case of the harmonic oscillator, let us select the coordinate system in Figure 8-3 with the origin at the equilibrium position of the particle and with the y-axis along the spring. According to Hooke's law, if the particle is displaced a distance y, the force on it is $-ky$, where the minus sign reflects the fact that the force is always directed toward the origin. To obtain the potential energy $V(y)$ for this case, it is necessary to find a function whose negative derivative, in accordance with (8-2), yields this force $-ky$. Since the derivative of $ky^2/2$ is ky, it follows that, in this case,

$$V(y) = \frac{1}{2}ky^2 \tag{8-7}$$

provided that the arbitrary constant is selected, as is customary, so that V vanishes at the equilibrium point. For example, when the particle is at its maximum displacement A, the potential energy associated with this position is $kA^2/2$, according to this formula.

According to Newton's law of gravitation, the force of attraction $F(y)$ on a particle of mass m at a distance y from the center of a planet or star of

mass M is

$$F(y) = -\frac{GMm}{y^2}$$

where the minus sign here is a consequence of our choice of orienting the y-axis radially outward (see Figure 8-4). Since $d(1/y)/dy = -1/y^2$, it follows, from (8-2), that

$$V(y) = -\frac{GMm}{y} \tag{8-8}$$

where this time the arbitrary constant has been selected so that when the particle is very far away from the planet its potential energy vanishes. This is the choice usually made.

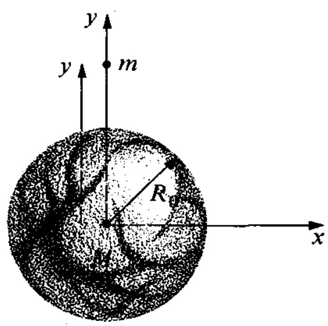


Figure 8-4

It should be emphasized that in writing down the potential energies in (8-6) through (8-8), the choices made for the arbitrary constants have no particular physical significance; however, in each case the choice actually made was the conventional one.

Example 8-3 Calculate the work W carried out by the gravitational field of a planet of mass M and radius R_0, if a small body of mass m is taken from the planetary surface to a distance $R (> R_0)$ from its center. Assume that all motion takes place along the radial direction.

Solution According to (8-1),

$$W = V(R_0) - V(R)$$

where R_0 and R are the initial and final distances of the body from the center of the planet. Making use of (8-8), we find by substitution into (8-1) that

$$\begin{aligned} W &= -\frac{GMm}{R_0} - \left(-\frac{GMm}{R}\right) \\ &= GMm\left(\frac{1}{R} - \frac{1}{R_0}\right) \end{aligned}$$

This work W is negative and represents the work carried out by the gravitational field of the planet as the body is raised through a vertical distance $(R - R_0)$. Consistent

with the result of Example 7-9, W is the negative of the work required of an external agent to raise the body through the same vertical distance.

Example 8-4 How much work is required to move a body of mass 0.5 kg from an elevation of 1 meter to an elevation of 10 meters:

(a) By the uniform gravitational field of the earth?

(b) By an external agent?

Solution As in Example 8-3, let us calculate the work by making use of (8-1).

(a) The potential energy V_1 of the body at the lower point is, according to (8-6),

$$\begin{aligned} V_1 &= mgy_1 = 0.5\ \text{kg} \times 9.8\,\frac{\text{m}}{\text{s}^2} \times 1\ \text{m} \\ &= 4.9\ \text{J} \end{aligned}$$

Correspondingly, at the higher elevation of 10 meters the potential energy V_2 is

$$\begin{aligned} V_2 &= mgy_2 = 0.5\ \text{kg} \times 9.8\,\frac{\text{m}}{\text{s}^2} \times 10\ \text{m} \\ &= 49\ \text{J} \end{aligned}$$

According to (8-1), the work W required to raise the body is the difference $(V_1 - V_2)$, and thus

$$\begin{aligned} W &= V_1 - V_2 \\ &= 4.9\ \text{J} - 49\ \text{J} = -44\ \text{J} \end{aligned}$$

The minus sign reflects the fact that the gravitational field of the earth is directed downward and thus carries out negative work on the body as it is raised upward.

(b) The work required of an agent other than the gravitational field to raise the body through the same height is the negative of this quantity, namely $(V_2 - V_1)$, and has the value $+44$ J.

8-4 The law of energy conservation

Consider a particle of mass m traveling along a straight line under the influence of a conservative force field $F(x)$. We define the *energy* E of the particle to be the sum of its kinetic and potential energies:

$$\boxed{E = \frac{mv^2}{2} + V(x)} \tag{8-9}$$

where v is the velocity of the particle when it is at the point x and $V(x)$ is the potential function associated with the given force $F(x)$. The main purpose of this section is to show that E as defined here is a constant of the motion. With this established it will follow that as the particle moves along its trajectory, the variation in its velocity v and position x is constrained by the condition that the quantity E as defined in (8-9) is constant in time. In particular, as the particle moves into regions of decreasing values for $V(x)$, it will speed up; conversely, it will slow down as it enters a region of

increasing values for $V(x)$. This constancy in time of the total energy E is the special form of the law of conservation of energy applicable for a particle confined to motion along a straight line. The fact that the energy of a physical system is the sum of its kinetic and potential energies is, however, very generally true.

To establish the constancy of the quantity E in (8-9), suppose that at the initial instant $t = 0$ the particle is at x_0 and has the velocity v_0. As a result of the action of the force $F(x)$, the particle will accelerate, and suppose that at a subsequent instant it is at the point x and has the velocity v. According to the definition of potential energy, the work W carried out by the force field during this time interval is the difference $[V(x_0) - V(x)]$. Moreover, according to the work-energy theorem, W can also be expressed as the difference $(\frac{1}{2}mv^2 - \frac{1}{2}mv_0^2)$ of the kinetic energy of the particle at these two times. Hence

$$W = V(x_0) - V(x) = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$$

or, rearranging, this may be expressed in the form

$$\frac{1}{2}mv^2 + V(x) = \frac{1}{2}mv_0^2 + V(x_0) \tag{8-10}$$

But the left-hand side is the sum of the kinetic and potential energies of the particle at the instant t, whereas the right-hand side is the corresponding sum at the initial instant $t = 0$. The constancy in time of the total energy E in (8-9) is thereby established.

Specialized forms of the result in (8-10) have actually come up previously in a number of instances. Thus, the substitution into (8-10) of the potential energy for the inverse-square field in (8-8) yields (5-11), and similarly the substitution of (8-7) yields the energy integral for the harmonic oscillator in (6-15). The result in (8-10) may thus be viewed as the generalization of these and other special cases.

It should be emphasized that the above derivation of the law of energy conservation applies only to the case of a single particle confined to motion along a straight line. It will become evident as we go along, however, that this law is valid for a much wider range of phenomena and physical systems than is implied by this derivation. Indeed, the importance of the law of energy conservation is due in large measure to the all-pervasiveness of its applicability. It is applicable even for physical situations for which Newton's laws themselves break down and are superseded by those of relativity and quantum mechanics.

Example 8-5 A body of mass 0.05 kg is thrown vertically upward with an initial speed of 10 m/s. To what maximum height does it rise?

Solution If the zero of potential energy is selected so that initially the potential energy of the particle vanishes, it follows that initially the total energy E of the

particle is purely kinetic. Hence E has the value

$$E = \frac{1}{2} m v_0^2 = \frac{1}{2} \times 0.05 \text{ kg} \times (10 \text{ m/s})^2$$
$$= 2.5 \text{ J}$$

On the other hand, at the peak of the particle's trajectory the velocity vanishes. Hence at this point its total energy must be purely potential and, according to (8-6), must have the value mgh, where h is the unknown height. According to the conservation of energy law the initial energy must be the same as that when the particle is at its maximum height. Thus

$$mgh = 2.5 \text{ J}$$

and this leads to the value $h = 5.1$ m. This same result could also have been achieved by direct usage of the formulas in Chapter 2.

Example 8-6 By differentiating (8-9), show explicitly that the total energy E is constant in time.

Solution According to the chain rule, the time derivative of $mv^2/2$ is $mv\, dv/dt$, while the time derivative of $V(x)$ is

$$\left(\frac{dV}{dx}\right)\left(\frac{dx}{dt}\right)$$

Hence, differentiating (8-9), we obtain

$$\frac{dE}{dt} = \frac{d}{dt}\left(\frac{1}{2} m v^2 + V(x)\right) = mv\frac{dv}{dt} + \left(\frac{dV}{dx}\right)\left(\frac{dx}{dt}\right)$$
$$= v\left[m\frac{dv}{dt} - F(x)\right]$$
$$= 0$$

where the third equality follows from (8-2) and the last from the second law. The constancy in time of E is thereby established.

8-5 Two- and three-dimensional motion

We shall now generalize the above ideas on potential energy and energy conservation to the case of a body accelerating along an arbitrary path through three-dimensional space.

Consider, in Figure 8-5, a particle of mass m moving along a curve C_1 from P_1 to P_2 under the action of a certain force $\mathbf{F}$. Let W represent the work carried out by the force $\mathbf{F}$ as the particle follows its trajectory from P_1 to P_2. We shall say that $\mathbf{F}$ is *conservative* provided there exists a function $V \equiv V(x, y, z)$, which depends only on the spatial coordinates (x, y, z) of the particle and is such that W is the difference

$$W = V_1 - V_2 \qquad \textbf{(8-11)}$$

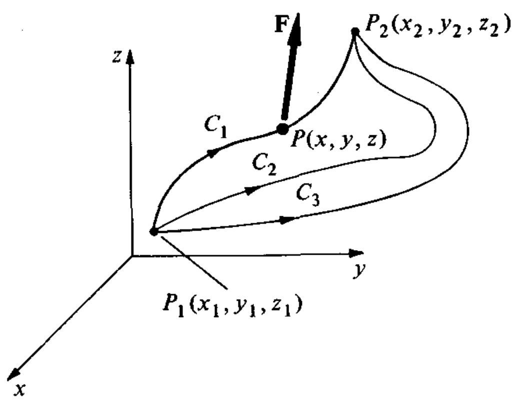


Figure 8-5

where V_1 is the value of V at the point P_1 with coordinates (x_1, y_1, z_1) and similarly for V_2. As in the analogous one-dimensional case, the quantity V is known as the *potential energy* function and characterizes an attribute of the body known as its potential energy. We emphasize that in order for the force field $\mathbf{F}$ to be conservative, it is necessary that (8-11) be valid for all points P_1 and P_2. In other words, if $\mathbf{F}$ is conservative, the work W must be the difference in (8-11) for any initial and final positions of the particle.

As in the corresponding one-dimensional case, no force field $\mathbf{F}$ that depends explicitly on the time t or on the velocity $\mathbf{v}$ of the body can be conservative. That is, there does not exist, in general, a potential function V as defined in (8-11) for such explicitly time- or velocity-dependent forces. However, unlike the one-dimensional case, not all force functions $\mathbf{F}$ that are independent of $\mathbf{v}$ and t are conservative. And this feature is *the* important distinction between one-dimensional force fields and others.

To clarify this distinction, let us consider a force field $\mathbf{F}$ which varies only with the position $\mathbf{r}$ of a particle. Assuming that $\mathbf{F}$ is conservative, it follows by the use of (7-20) that the definition in (8-11) may be expressed in the form

$$\int_{P_1}^{P_2} \mathbf{F} \cdot d\mathbf{r} = V_1 - V_2 \tag{8-12}$$

Now, in general, the line integral in the left-hand side depends on the particular path followed by the particle as it goes from P_1 to P_2. That is, since $\mathbf{F}$ generally varies from point to point, the integral in (8-12) will also in general be different for different paths joining P_1 to P_2. But on the other hand, the right-hand side of (8-12) is the difference of the potential function evaluated at P_1 and P_2 and thus depends only on P_1 and P_2 and *not* on the path connecting them. Thus follows the important characterization of a conservative force $\mathbf{F}$ as one whose *line integral between any two points is the same for all paths connecting these two points.* In Figure 8-5, for example, if $\mathbf{F}$ is conservative, then the line integral of $\mathbf{F}$ along C_1 is the same as along C_2 as well as along *all* other paths, such as C_3, joining P_1 and P_2. In one dimension, on the other hand, there is only one path connecting any two

points, so this ambiguity does not arise. It is for this reason that all one-dimensional force functions of the form $F(x)$ are conservative.

An equivalent way of characterizing a conservative force field can be obtained by use of the fact that the line integral in (8-12) is independent of the path joining P_1 and P_2. For this implies that the integrand $\mathbf{F} \cdot d\mathbf{r}$ in this equation must be a perfect differential. Thus follows the differential form

$$dV = -\mathbf{F} \cdot d\mathbf{r} \tag{8-13}$$

whose validity can also be ascertained by direct substitution into (8-12). The significance of (8-13) is illustrated in Figure 8-6 which shows the relation between the values V and $(V + dV)$ of the potential energy function at two points at a relative infinitesimal displacement $d\mathbf{r}$.

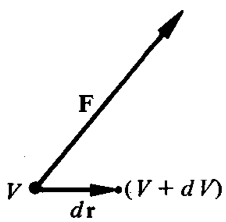


Figure 8-6

With regard to the law of conservation of energy, the situation in three dimensions differs very little from that in the simpler one-dimensional framework. Equating the work W in (8-11) carried out by a conservative force to the difference in kinetic energies as given by the work-energy theorem we find that, just as for the one-dimensional case, the total energy E defined by

$$\boxed{E = \frac{m\mathbf{v}^2}{2} + V} \tag{8-14}$$

is constant in time. Here, however, $\mathbf{v}$ is a vector, and the potential energy V of the particle depends, in general, on the three coordinates (x, y, z) of the particle. Hence the specification of the position of the particle does not yield its velocity $\mathbf{v}$, but only its speed v. Unlike the one-dimensional case, then, the law of conservation of energy does not in general yield a complete particle trajectory for the case of three-dimensional motion.

Example 8-7 Show that the potential energy $V(r)$ of a particle of mass m at a distance $r(>R)$ from a planet of mass M and radius R is

$$V = -\frac{GMm}{r} \tag{8-15}$$

Solution Let us calculate the work W carried out in taking the particle from a reference point P_1 a distance r_1 from the center of the planet to an arbitrary point P_2 at a distance r_2 (see Figure 8-7). The force $\mathbf{F}$ on the particle when it is at a position $\mathbf{r}$

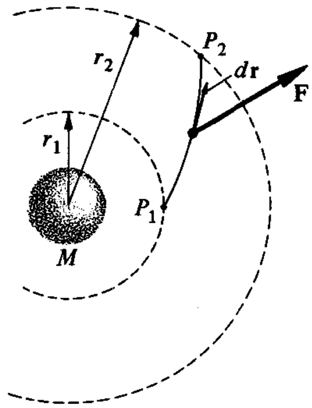


Figure 8-7

relative to the center of the planet is

$$\mathbf{F} = -\frac{GMm}{r^3}\mathbf{r}$$

since $(-\mathbf{r}/r)$ is a unit vector directed radially inward. Hence when the particle undergoes the infinitesimal displacement $d\mathbf{r}$, the work dW carried out is

$$dW = \mathbf{F} \cdot d\mathbf{r} = -\frac{GMm}{r^2}\,dr$$

where dr is the component of $d\mathbf{r}$ along the radial direction. Integrating, we find, by use of (7-20), that

$$W = \int_{P_1}^{P_2} \mathbf{F} \cdot d\mathbf{r} = -\int_{r_1}^{r_2} \frac{GMm}{r^2}\,dr = \frac{GMm}{r}\bigg|_{r_1}^{r_2}$$
$$= \frac{GMm}{r_2} - \frac{GMm}{r_1}$$

The validity of (8-15) therefore follows by comparison with (8-12).

It is interesting to note the similarity between (8-15) and its "one-dimensional version" in (8-8). In precisely the same way as shown here, it can be established that the potential energy associated with the one-dimensional motion of a particle in a uniform field in (8-6) is also applicable if the particle has a horizontal component of velocity.

Example 8-8 A satellite of mass m orbits about a planet of mass M in a circular orbit of radius a.

(a) What is the speed v of the satellite?

(b) Show that the total energy of the satellite is numerically equal to one half of its potential energy.

Solution

(a) Making use of Newton's laws of motion and of universal gravitation, we have

$$\frac{GMm}{a^2} = \frac{mv^2}{a}$$

since the satellite has only the centripetal acceleration v^2/a. Solving for v^2, we obtain the sought-for value

$$v^2 = \frac{GM}{a}$$

(b) According to (8-14) and (8-15), the energy E of the satellite is

$$E = \frac{1}{2}mv^2 - \frac{GMm}{a}$$

Hence, substituting for v^2 by using the result of (a), we obtain

$$\begin{aligned} E &= \frac{1}{2}m\left(\frac{GM}{a}\right) - \frac{GMm}{a} \\ &= -\frac{1}{2}\frac{GMm}{a} \end{aligned}$$

which is the sought-for result.

8-6 Motion under gravity

One of the very important applications of the principle of the conservation of energy is to physical situations involving the motion of bodies in the gravitational field of the earth. In the present section we shall discuss some problems of this type.

Consider first the case of a satellite or other body of mass m orbiting about the earth. According to (8-15), the potential energy V of the satellite is $-GMm/r$, with M the mass of the earth and r the distance to the satellite from the earth's center (see Figure 8-8). In terms of the velocity $\mathbf{v}$ of the satellite at an instant when its position relative to the center of the earth is $\mathbf{r}$, the law of conservation of energy has the form

$$E = \frac{1}{2}m\mathbf{v}^2 - \frac{GMm}{r} \tag{8-16}$$

with E the constant energy of the satellite. Hence if the velocity $\mathbf{v}_0$ of the

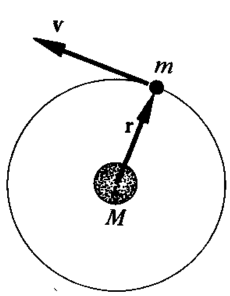


Figure 8-8

satellite is known at an instant when its position is $\mathbf{r}_0$, then (8-16) determines a value for the constant E/m in terms of these known quantities. With this knowledge, the magnitude of the velocity $\mathbf{v}$ at any other position of the satellite in its orbit may be determined uniquely. Note the significant feature that it is not necessary to know the mass of the satellite in order to make this prediction. Indeed, it is not possible to determine the mass of a satellite by measuring only its orbital parameters; the mass m invariably cancels out of the equations of motion.

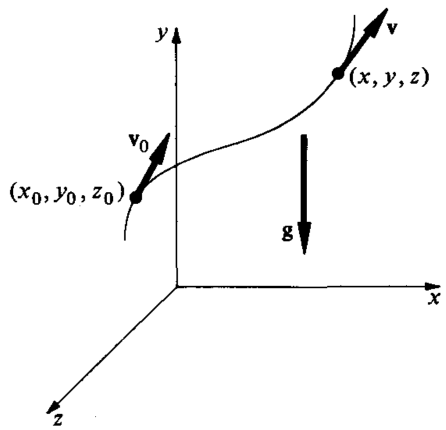


Figure 8-9

As a second application let us consider motion in a uniform gravitational field $\mathbf{g}$. The potential energy of a body of mass m is in this case given by mgy, provided that the coordinate system is as defined in Figure 8-9. Substitution into (8-14) leads to

$$E = \frac{1}{2}m\mathbf{v}^2 + mgy \tag{8-17}$$

with E the constant energy of the body. If initially the particle is at the point with coordinates (x_0, y_0, z_0) where it has a velocity $\mathbf{v}_0$, then

$$\frac{1}{2}mv^2 + mgy = \frac{1}{2}mv_0^2 + mgy_0$$

with v the magnitude of velocity $\mathbf{v}$ at the subsequent instant when the coordinates of the particle are (x, y, z) (see Figure 8-9). Dividing throughout by the factor $m/2$, we obtain the familiar form

$$v^2 = v_0^2 - 2g(y - y_0) \tag{8-18}$$

a relation that can also be obtained by taking the sum of the squares of (3-27) and eliminating the variable t by use of (3-26). Also of interest is the fact that (8-18) is formally identical to the corresponding one-dimensional formula in (2-18). However, in that case we dealt with motion in one dimension, so v represented the velocity of the particle along the x-direction. In the present case, $\mathbf{v}$ is a vector with components along the x-, y-, and z-axes, so here v

represents the magnitude of this velocity. The fact that (8-18) does not supply complete information about the motion of the system is apparent, since the particle's x- and z-coordinates do not enter into this relation at all.

Under special circumstances it is also possible to use the conservation-of-energy principle in (8-17) to study the motion of a particle under the combined influence of gravity and certain other, not necessarily conservative, forces. This will be the case, for example, if the trajectory of the particle is such that *these additional forces do no work on it throughout its motion.* For in that case the work carried out by such, not necessarily conservative, forces will vanish and thus (8-12) will be valid, with **F** the total force acting on the particle. In other words, (8-18) will be applicable, provided that the direction of any nongravitational forces are always at right angles to the velocity of the particle throughout its motion.

A typical case is illustrated in Figure 8-10 which shows a body of mass m starting out at a velocity $\mathbf{v}_0$ from the top of a *smooth*, inclined plane of angle α. Because the plane is smooth it follows that the force which the plane exerts on the body does no work on it as it slides down the plane. Accordingly, (8-18) may be used *as if this reaction force were not present.* Assuming, as shown in Figure 8-10, that the body slides the full distance $h/\sin\,\alpha$ down the plane so that it drops through a vertical distance h, it follows from the conservation-of-energy principle that

$$\frac{1}{2}mv_0^2 + mgh = \frac{1}{2}mv^2 + 0$$

where the potential energy has been taken to be zero at the bottom of the plane. Solving for v^2, we regain the applicable form of (8-18).

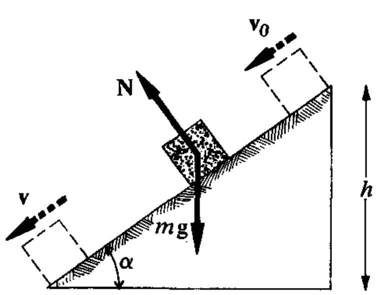


Figure 8-10

Brief reflection of this derivation shows that the fact that the surface is flat is not crucial. What is important is that the surface be smooth so that the reaction force exerted by the surface on the body carries out no work during the body's descent. Figure 8-11 shows identical particles sliding down three smooth surfaces, each of a different shape. Assuming that the particles start out at rest at the top of the various surfaces at points A, A', and A'', respectively, each of which is at the same elevation h above the ground, it follows that since the surfaces are smooth they carry out no work on the

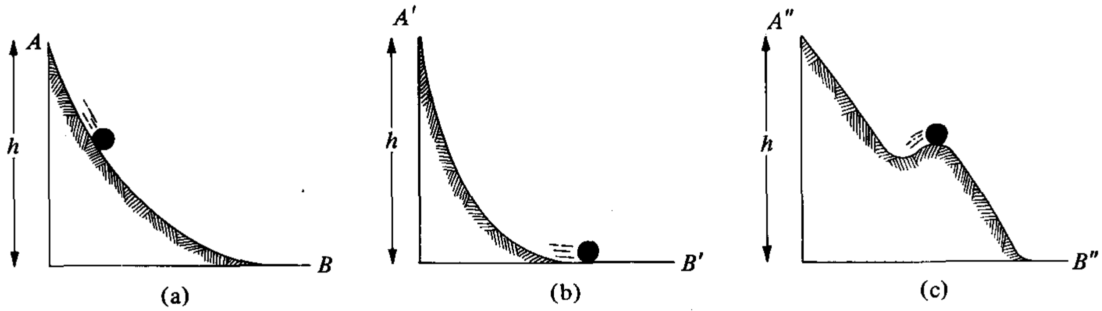


Figure 8-11

particles. Thus the speed of each particle when it reaches the bottom of its surface is as if that particle had dropped through the vertical height h. In other words, using the conservation-of-energy principle and the fact that each of the particles starts out at rest at the same elevation h, we obtain

$$0 + mgh = \frac{1}{2}m\mathbf{v}^2 + 0$$

where $\mathbf{v}$ is the velocity at the bottom (at points B, B', and B'', respectively). Thus, consistent with (8-18), we see that $\mathbf{v}^2 = 2gh$ in each case, *independent of the shape of the surface.* Note that even though the terminal speeds are in each case the same, the times of descent are, in general, different; these times cannot be obtained by exclusive use of the law of energy conservation.

Example 8-9 A particle of mass m is released at rest at the top of a smooth, inclined plane of angle α and height h at the bottom of which is a smooth circular hoop of radius a (see Figure 8-12).

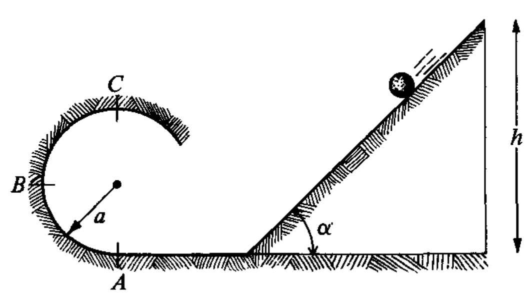


Figure 8-12

(a) What is the velocity of the particle at point A?
(b) Assuming that the particle gets to point C, calculate its velocity there.
(c) Calculate the reaction force on the particle at point B.

Solution

(a) Since all contacts are smooth, the conservation-of-energy principle, with the potential energy associated with the uniform gravitational field of the earth, is applicable. Accordingly, if $\mathbf{v}_A$ is the velocity of the particle when it reaches point A,

then assuming the zero of potential at A we find

$$0 + mgh = \frac{1}{2} m\mathbf{v}_A^2 + 0$$

and this leads to

$$v_A = \sqrt{2gh}$$

(b) Similarly, assuming that $h > 2a$, the speed of the particle when it reaches point C is

$$v_C = \sqrt{2g(h - 2a)}$$

(c) At point B the speed $|\mathbf{v}_B|$ of the particle is found as in the above arguments to be given by the formula $v_B = \sqrt{2g(h-a)}$. Furthermore, at this point the normal force $\mathbf{N}$, which is always directed radially inward, must produce the required centripetal acceleration, v_B^2/a. Accordingly, an application of the second law leads to

$$N = m\frac{v_B^2}{a} = \frac{2\,mg}{a}(h - a)$$

where the second equality follows by use of the above formula for v_B.

Example 8-10 Consider, in Figure 8-13, a pendulum of length l. Assuming that the bob is initially displaced an angle θ_0 and then released at rest, calculate the speed v of the bob as a function of the angular displacement θ.

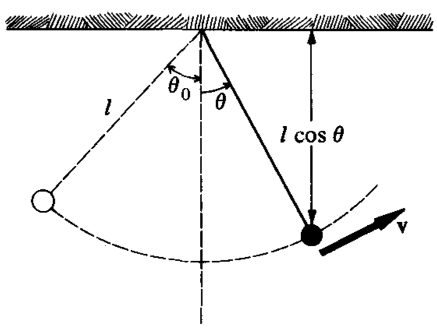


Figure 8-13

Solution As the bob moves along its circular orbit, it experiences two forces: (a) the force of gravity $m\mathbf{g}$, and (b) the tension $\mathbf{P}$ in the supporting string. But the latter is perpendicular to the velocity $\mathbf{v}$ of the bob throughout the motion. Hence $\mathbf{P}$ carries out no work and we may use the ideas of energy conservation, assuming that the bob has the potential energy in (8-6).

If we select the zero potential to be at the lowest point in the orbit, it follows that the initial potential energy of the bob is $mgl(1 - \cos\theta_0)$. Correspondingly, when it is at the angular displacement θ, its potential energy V is $mgl(1 - \cos\theta)$. Since initially the kinetic energy vanishes, it follows from the conservation-of-energy law that

$$0 + mgl(1 - \cos\theta_0) = \frac{1}{2} m\mathbf{v}^2 + mgl(1 - \cos\theta)$$

where v is the velocity of the bob when it is at the angle θ. Solving for v^2, we obtain

$$v^2 = 2gl(\cos\theta - \cos\theta_0)$$

and this determines v for any angle θ. Consistent with the initial condition we find that $v^2 = 0$ at $\theta = \theta_0$. The velocity v is a maximum at the lowest point of the orbit, where $v^2 = 2gl(1 - \cos\theta_0)$.

Example 8-11 Consider, in Figure 8-14, a pendulum of length l suspended a distance $(l - l_1)$ vertically above a small peg C. Suppose that the bob is initially displaced by an angle θ_0 and then released at rest. Calculate the velocity v of the bob at the instant shown in the figure when it is traveling in a circular orbit of radius l_1 and has an angular displacement ϕ with respect to the vertical. (See Section 7-2.)

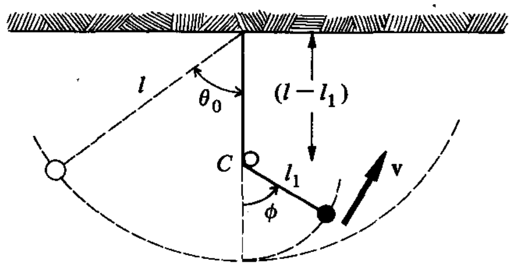


Figure 8-14

Solution As in the Example 8-10, the tension P that the string exerts on the bob carries out no work on the bob throughout its motion. Hence, the law of energy conservation for motion in a uniform gravitational field is again applicable. Taking the zero point of potential energy at the lowest point in the orbit, the potential energy of the bob at the instant shown is $mgl_1(1 - \cos\phi)$. Hence, since it starts out at rest and its initial potential energy is $mgl(1 - \cos\theta_0)$, it follows that

$$0 + mgl(1 - \cos\theta_0) = \frac{1}{2}mv^2 + mgl_1(1 - \cos\phi)$$

and this leads to

$$v^2 = 2g[l(1 - \cos\theta_0) - l_1(1 - \cos\phi)]$$

It is left as an exercise to confirm that, consistent with Galileo's observations, the speed v of the bob vanishes at an angle ϕ_m at which its vertical distance below the point of suspension is precisely the same as was its initial value $l\cos\theta_0$.

8-7 Bound states

Two or more bodies are said to be in a *bound state* relative to each other if their motions are such that they remain in each other's proximity for extended periods of time. One example of a bound state is the earth-moon system; another is the hydrogen atom, which is a bound state of an electron and a proton; a third is a diatomic molecule. The purpose of this section is to

show how by use of the energy-conservation principle we may analyze certain physical systems such as these for the existence of bound states. To simplify matters, let us confine ourselves to the case of particles undergoing one-dimensional motion.

An interaction that is often used to describe the behavior of a diatomic molecule is the *Lennard–Jones* 6-12 *potential.* In terms of the separation distance x between the two constituent atoms this potential has the form

$$V(x) = V_0\left[\left(\frac{x_0}{x}\right)^{12} - 2\left(\frac{x_0}{x}\right)^{6}\right] \qquad (x>0) \tag{8-19}$$

where V_0 and x_0 are two positive parameters. The force $F(x)$ associated with this potential may be calculated by use of (8-2) to be

$$F(x) = \frac{12V_0}{x_0}\left[\left(\frac{x_0}{x}\right)^{13} - \left(\frac{x_0}{x}\right)^{7}\right]$$

and Figure 8-15 shows a plot of this potential and the associated force as a function of the interatomic spacing x. Note from these graphs that at very short distances ($x \ll x_0$) the potential energy is positive and large, and since the curve for $V(x)$ has a negative slope in this region it follows that the force is here repulsive. For large separations, on the other hand, the potential is weak and negative, and since its slope is very small and positive it follows from (8-2) that in this outer region the force is attractive and weak. In brief, for distances very much greater than the parameter x_0 the potential function in (8-19) corresponds to a weak attractive force, whereas for distances very much smaller than this it is strong and repulsive. These features are characteristic of both interatomic and internuclear forces, and it is for this reason that potentials of the form in (8-19) are often used to describe the dynamics of these two classes of physical systems.

Consider now the motion of a particle in the force field associated with the potential in (8-19). Suppose that the initial conditions are such that the total

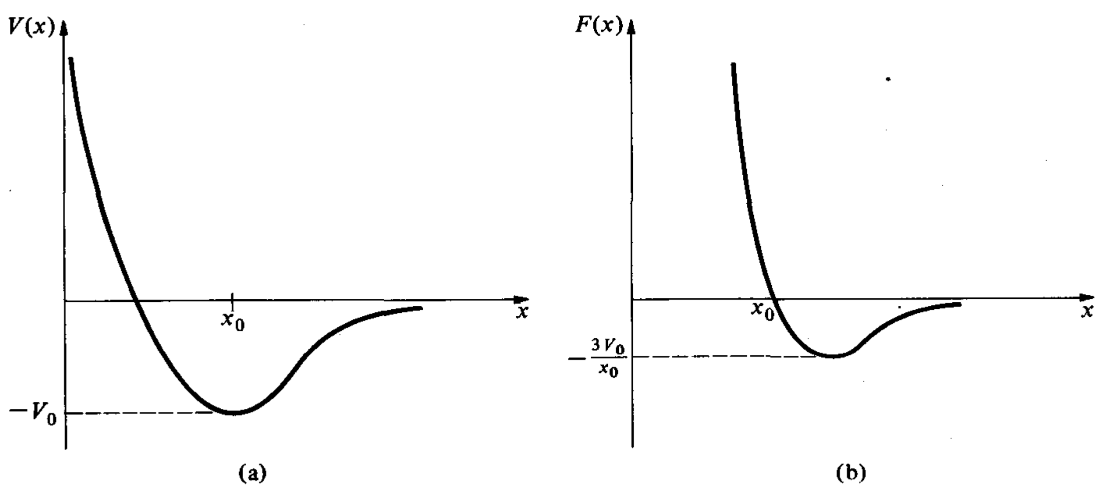


Figure 8-15

energy of the particle E_1 is negative.[1] This means that the quantity E_1 given by

$$E_1 = \frac{1}{2}mv^2 + V_0\left[\left(\frac{x_0}{x}\right)^{12} - 2\left(\frac{x_0}{x}\right)^6\right] \tag{8-20}$$

is negative. To analyze the motion consider, in Figure 8-16, a plot of $V(x)$ as a function of x, and draw on this graph a line at a distance $-E_1$ below and parallel to the x-axis. Let us use the symbols x_1 and x_2 for the x-coordinates of the two points at which this horizontal line intercepts the potential curve.

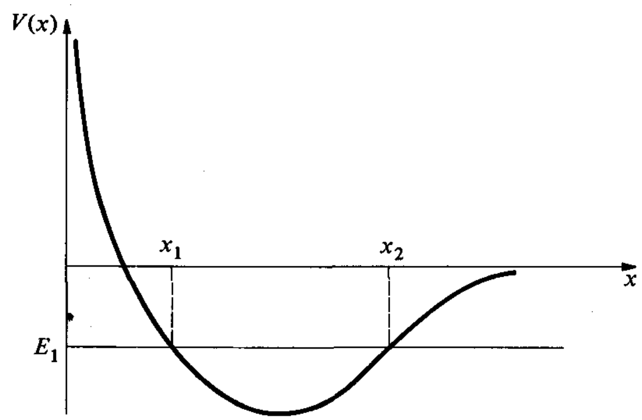


Figure 8-16

Referring back to (8-20), we see that since at x_1 and x_2 the potential energy is equal to the total energy E_1, at these points the velocity of the particle must vanish. Moreover, in the regions $0 < x < x_1$ and $x > x_2$ the potential energy $V(x)$ is *greater* than the total energy E_1, and since according to (8-20) there are no real values for the velocity v that will satisfy this relation, it follows that the particle can never enter this region. In other words, if the energy of the particle has the negative value E_1, then its motion is confined to those values of x that satisfy the inequality $x_1 \leq x \leq x_2$, and in this case a bound state exists. More generally, since the minimum value of $V(x)$ is $-V_0$, it follows that if the initial conditions are such that the total energy E is greater than $-V_0$ but still negative, then the particle will be in a bound state. The total energy E can never be less than $-V_0$, since there are no values for v that will satisfy (8-20) in this case.

By the use of arguments of this type a considerable amount of information can be obtained about the motion of the particle. Assuming that the situation is as described in Figure 8-16, suppose that initially the particle is at rest at x_1. Since there it experiences a force directed to the right, it will start to move in the direction of increasing values for x. In principle, its velocity v at

[1]Implicit in (8-19) is the choice of the additive constant so that zero potential energy corresponds to infinite separation. The negative values for V in Figure 8-15a are a consequence of this choice. With this selection for the zero point of energy, a negative value for E_1 means therefore that energy in the amount $-E_1$ is required to break up the bound system and to bring the particles to a state of rest at infinite separation.

any such point can be calculated by solving (8-20) for v and inserting the known values of E_1 and x. Eventually the particle arrives at the point x_2, where again it has zero velocity. Here it experiences an attractive force that now causes it to accelerate backward toward x_1. Eventually it arrives back at x_1, where it will be in its original situation at rest. It follows, therefore, that the particle moves back and forth between the points x_1 and x_2 with periodic motion. And even though it would be very difficult to integrate Newton's laws of motion in order to find the precise trajectory of the particle, by use of the energy integral in (8-20) we see that a fairly detailed picture of its motion can be obtained.

We emphasize that the above analysis is applicable only if the initial conditions are such that the total energy is negative. It can easily be established that if the total energy is positive, then no bound state can exist. For, in this case, if in Figure 8-16 a horizontal line is drawn above the x-axis, then even though there is a smallest positive value for x below which the particle can never go, there is no upper limit to the x-coordinate of the particle's position. In other words, for positive values for E, even though there is a shortest distance between the particles, there is no upper limit to their separation distance and thus they are not in a bound state.

Example 8-12 For a given value for E_1 satisfying the relation $-V_0 < E_1 < 0$, so that a bound state exists, find the values for x_1 and x_2 in Figure 8-16 that determine the extreme values for the position of the particle.

Solution At the points x_1 and x_2 the velocity v vanishes. Setting v in (8-20) to zero, we obtain a quadratic equation for the variable $z = (x_0/x)^6$:

$$z^2 - 2z - \frac{E_1}{V_0} = 0$$

This is readily solved, and the roots are

$$z = 1 \pm \sqrt{1 + E_1/V_0}$$

Note that since E_1 is negative but always smaller in magnitude than V_0, the radical is bounded between 0 and 1. Making use of the above definition for z, we find for x_1 and x_2 the values

$$x_{1,2} = x_0\{1 \pm (1 + E_1/V_0)^{1/2}\}^{-1/6}$$

†8-8 Nonconservative forces

If the force acting on a particle depends explicitly on $\mathbf{v}$ or on t, then since a potential energy function cannot be associated with such a force, it follows that the conservation-of-energy theorem simply does not apply. Nevertheless, even for these situations the concept of energy is often very useful, and the purpose of this section is to describe what meaning the term "energy" can have in this context.

Consider a particle of mass m confined to one-dimensional motion in a force field that consists of the sum of a conservative and a nonconservative part. Let $V(x)$ represent the potential energy function associated with the former, and suppose that the nonconservative force F_v is a viscous force of the form $-bv$ in (5-18). The equation of motion of the particle is

$$m\frac{dv}{dt}=-\frac{dV}{dx}-bv \tag{8-21}$$

where use has been made of (8-2). Let us recall that the parameter b is a positive constant and that the negative sign in front of the viscous-force term reflects the fact that the viscous force always opposes the motion of the particle.

In order to explore the possibility of assigning an energy to this particle, let us define a quantity $\mathscr{E}$ to be the sum of the particle's kinetic energy and potential energy; that is

$$\mathscr{E}=\frac{1}{2}mv^2+V(x) \tag{8-22}$$

so that $\mathscr{E}$ would be the constant energy of the particle if the viscous force were not present. However, the viscous force does exist and does have an influence on the motion of the particle. In order to see the effect of this force let us differentiate (8-22). Making use of the chain rule in Appendix A, we obtain

$$\begin{aligned}\frac{d\mathscr{E}}{dt}&=\frac{d}{dt}\left[\frac{1}{2}mv^2+V(x)\right]\\&=mv\frac{dv}{dt}+\frac{dV}{dx}\frac{dx}{dt}\\&=v\left[m\frac{dv}{dt}+\frac{dV}{dx}\right]\end{aligned}$$

where the final equality follows since $v=dx/dt$. Substituting now for the square bracket in the last equality, we find by use of the equation of motion, (8-21), the result

$$\frac{d\mathscr{E}}{dt}=-bv^2 \tag{8-23}$$

so that as long as the particle travels it will *lose* energy at the rate bv^2. Thus if a particle moves under the combined action of a conservative force $-dV/dx$ and the viscous force $F_v=-bv$, then its total energy $\mathscr{E}$, given in the conventional way to be the sum of the kinetic and potential energy in (8-22), is *not* conserved! But rather it decreases monotonically at a rate given in (8-23). It is for this reason that viscous forces and other nonconservative forces that cause the energy of a particle to decrease are also called *dissipative* forces.

A simple physical interpretation can be given to the term $-bv^2$, which appears on the right-hand side of (8-23), in the following way. The rate dW_v/dt at which the viscous force F_v carries out work on the particle is vF_v, according to (7-26) and (7-27). Therefore, since $F_v = -bv$, it follows that (8-23) may be expressed equivalently as

$$\frac{dW_v}{dt} = \frac{d\mathscr{E}}{dt} \tag{8-24}$$

and this states that the power, or the *rate* at which the viscous force F_v carries out work on the particle, is numerically equal to the rate of increase of the energy $\mathscr{E}$ of the particle. Or, equivalently, writing (8-24) in the form

$$\frac{d}{dt}(-W_v) = -\frac{d\mathscr{E}}{dt}$$

and noting that $(-W_v)$ represents the work carried out *on* the viscous medium by the particle, we may interpret (8-24) by the statement:

The rate at which the particle carries out work on the dissipative force field F_v is precisely the same as the rate at which the particle loses energy.

In this generalized sense then, the conservation-of-energy principle is valid for the larger system consisting of the particle under consideration plus the particles comprising the viscous medium. Physically, this means that on the one hand when a particle moves in a viscous medium it slows down and loses energy, but on the other this energy loss shows up in the form of heat of the medium. Provided then that the system of interest is defined to consist of the particle plus all matter that may be nearby and can affect the motion of the particle, energy is invariably conserved. However, if we are concerned only with the particle itself, then for the smaller system, because of the influence of the viscous term F_v, energy is not conserved.

In Problem 20 it will be established that if a body moves under the action of two forces, one of which is conservative while the other, $\mathbf{F}_v$, may or may not be, then

$$\frac{d}{dt}\left(\frac{1}{2}m\mathbf{v}^2 + V\right) = \mathbf{v} \cdot \mathbf{F}_v \tag{8-25}$$

where V is the potential energy associated with the conservative force. Thus, the time rate of change of the kinetic plus potential energy of the particle is the same as the rate at which the force $\mathbf{F}_v$ carries out work on it.

Example 8-13 Suppose that a body moves through a viscous medium for which the coefficient b in (8-21) has the value $b = 0.3$ kg/s. At what rate is the particle losing energy when it has a speed of 0.5 m/s?

Solution The rate of energy loss is the same as the negative of the rate at which the viscous force does work on the particle. The energy loss is thus

$$-\frac{d\mathscr{E}}{dt} = +bv^2$$

which, since $b = 0.3$ kg/s, becomes for $v = 0.5$ m/s

$$-\frac{d\mathscr{E}}{dt} = (0.3 \text{ kg/s}) \times (0.5 \text{ m/s})^2 = 7.5 \times 10^{-2} \text{ W}$$

Note that this result applies regardless of what other *conservative* forces may be acting on the particle.

8-9 Summary of important formulas

A force field $\mathbf{F}$ is conservative provided that there exists a potential energy function V which depends only on position in space and has the property that if a particle goes from a point P_1 to another point P_2, the work W carried out by $\mathbf{F}$ is

$$W = V_1 - V_2 \tag{8-11}$$

where V_1 is the value of V at P_1, and similarly for P_2. For the special case of one-dimensional motion, $V(x)$ and $F(x)$ are related by

$$F(x) = -\frac{dV}{dx} \tag{8-2}$$

or, equivalently,

$$V(x) = -\int_{x_0}^{x} F(x)\,dx \tag{8-3}$$

If a body of mass m travels under the action of a conservative force field, then its total energy E defined as the sum of its kinetic and potential energies

$$E = \frac{1}{2}m\mathbf{v}^2 + V \tag{8-14}$$

is constant in time. Hence, if $|\mathbf{v}|$ and V are known at any one instant, E may be evaluated and (8-14) can then be used to calculate the speed v of the particle in terms of its position at any other time.

QUESTIONS

1. Review briefly the meaning of the terms (**a**) conservative force field; (**b**) potential energy; (**c**) bound state; and (**d**) dissipative force.
2. If $\mathbf{a}$ is the acceleration of a particle of mass m under the action of the force $\mathbf{F}$, then the quantity $(\mathbf{F} - m\mathbf{a})$ does not vary in time. Why is this quantity nevertheless not a constant of the motion as we have defined it?

3. Give three examples of force fields that are conservative. Give three examples of force fields that are not conservative.
4. Some authors define a conservative force field as one that carries out no work on a particle when it is taken around a *closed* path. Is this consistent with our definition in (8-1) or (8-11)? Explain.
5. Explain why it is that a one-dimensional force field $F(x)$ which depends only on the single spatial coordinate x is invariably conservative. Why is this *not* true for a three-dimensional force field $\mathbf{F}$ which also depends only on position in space?
6. A body of mass m is taken along the path shown in Figure 8-17 from the point with the coordinates (x_1, y_1, z_1) to the point (x_2, y_2, z_2). Assuming that gravity acts vertically downward along the negative z-axis, how much work is carried out on the body by the gravitational field?

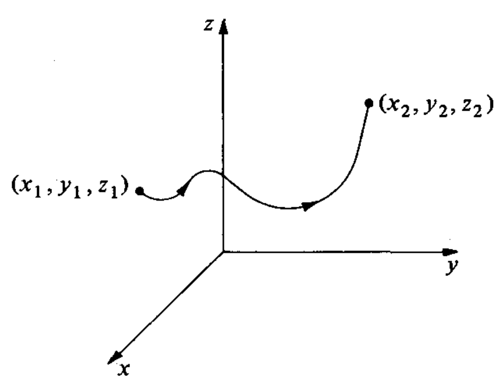


Figure 8-17

7. Figure 8-18 shows a block of mass m traveling to the right at the instantaneous velocity $\mathbf{v}$ along a rough, horizontal surface. If the coefficient of friction between the block and the surface is μ, then, as shown, there is a retarding frictional force $\mathbf{f}$ of magnitude μmg acting to the left along the negative x-axis. Explain why the

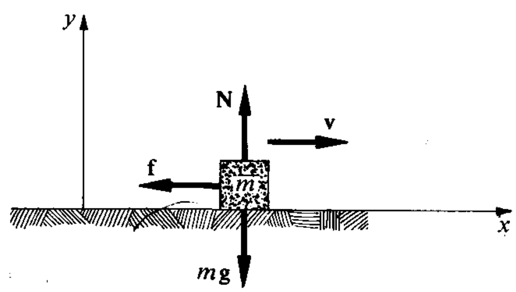


Figure 8-18

function μmgx, whose negative derivative yields the frictional force $\mathbf{f}$, *cannot* be the potential energy for the block. (*Hint*: Would this potential function work if the block were moving along the negative direction of the x-axis?)
8. Consider again the block moving along the rough, horizontal surface in Figure 8-18. What has happened to the total energy of the block while it is in the process of slowing down? For what total system would energy be conserved?
9. Explain what is meant by the statement: "On a microscopic scale energy is always conserved."
10. Explain how it can come about that, for certain force fields or combinations of force fields, for some particle trajectories energy will be conserved whereas for others it will not be conserved. Give an example of such a force field.
11. A ball rolls off the edge of a table and bounces on the floor below. Is it possible for the ball to achieve a height higher than the tabletop subsequently? Explain.
12. Figure 8-19 shows two bodies of mass m_1 and m_2 attached to a frictionless Atwood's machine of negligible inertia. Is the energy of either block conserved by itself? Explain.
13. Explain why the sum of the total energies kinetic plus potential of the blocks in the Atwood's machine in Figure 8-19 is conserved. (*Hint*:

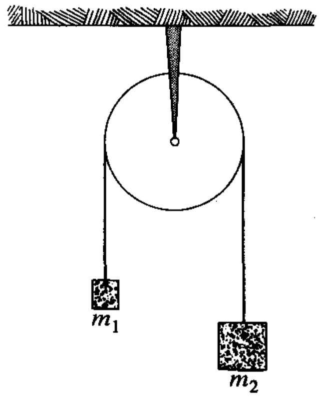


Figure 8-19

Why is the total work carried out by the tension in the string on both blocks zero?)

14. An experiment carried out on two different slides, as in Figure 8-11, shows that the particles reach the bottom in different times. Is this in contradiction with the conservation-of-energy principle? For which of the three slides shown in the figure would you expect this time to be least? Explain.
15. A satellite orbits about the earth. The point of its orbit at which it is farthest from the earth is known as the *apogee*, and the point at which it is closest to the center of the earth is known as its *perigee*. Explain why the satellite will be traveling faster when it is at perigee than when it is at apogee.
16. Show by use of the conservation-of-energy principle in the form given in (8-16) that when the satellite of a planet is at apogee then its speed is minimum, and that when it is at perigee its speed is maximum.
17. Why is it that even though the mass m' of the satellite appears explicitly in the conservation-of-energy statement, (8-16), we cannot use results predicted by this formula to deduce the mass of the satellite? Assume that the quantity GM is known and that v and r are easily measurable.
18. Suppose that a particle of mass m moves in a one-dimensional conservative force field associated with the potential function $V(x)$ shown in Figure 8-20. Assuming that initially the particle is at the point $x = 0$ and has a velocity v_0 directed to the right, will it speed up or slow down as it enters the region between x_1 and x_2 where $V(x)$ has the value $-V_0$?

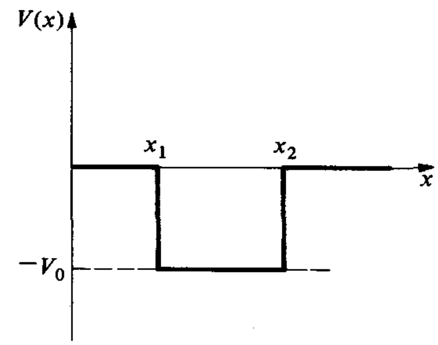


Figure 8-20

19. If a particle moves in the potential field $V(x)$ in Figure 8-20, is it possible for the total energy E of the particle to be less than the value $-V_0$? Is it ever possible for the total energy E of a particle to be less than the minimum of its potential energy?
20. Under what circumstances (if any) will a particle moving in the potential field shown in Figure 8-20 be in a bound state?
21. A particle of mass m undergoes vertical oscillations at the end of a hanging spring of constant k. Explain why its total energy $(\frac{1}{2}mv^2 + \frac{1}{2}kx^2)$, where x is its vertical displacement from equilibrium, is a constant of the motion even though as the particle moves up and down the gravitational field of the earth does work on it.
22. Under what circumstances is it possible to treat a fictitious force—arising because of our using a non-Newtonian reference frame—as being conservative?

PROBLEMS

1. What is the change in potential energy of a man of mass 70 kg who is in an elevator that rises a vertical distance of 15 meters? What would be the change in his potential energy if the elevator descended a distance of 15 meters?
2. A 200-gram baseball struck by a batter achieves a maximum height of 52 meters on its flight to the outfield. What is its change in potential energy during its flight assuming that it was struck when it was a distance of 2 meters above the ground? How much work was carried out on the ball by the force of gravity? Is the detailed path followed by the ball relevant?
3. If the bob of a pendulum of length 1.0 meter has a mass of 30 grams and if it is released at rest at an angle of 60° with respect to the vertical, what is the change in its potential energy when it reaches the bottom of its swing?
4. A uniform rod of mass m and length l is in a horizontal position and pivoted at one end (see Figure 8-21). If the rod is allowed to swing freely under the action of gravity until it reaches a vertical position as indicated by the dotted line in the figure:
   (a) Show that the change in potential energy, dV, of an element of length dr of the rod at a distance r from the pivot may be expressed as

$$dV = -\frac{mg}{l} r\, dr$$

   (b) Making use of the result of (a), calculate by integration the change in the potential energy of the total rod as it swings about the pivot point from the horizontal to the vertical position.

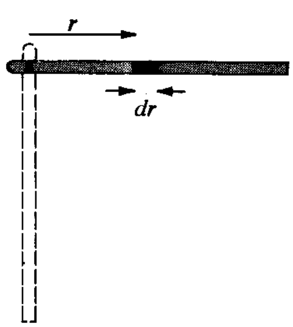


Figure 8-21

5. What is the change in potential energy of a 4.0×10^5-kg airplane as a result of its taking off from the ground and achieving an elevation of 10 km?
6. Consider a rocket of mass 10^5 kg that takes off from the surface of the earth at a point near the equator and achieves an altitude of one terrestial radius at a point directly above the earth's north pole. Using the values $GM = 4.0 \times 10^{14}$ N-m²/kg and $R_0 = 6.38 \times 10^6$ meters, where M and R_0 are the mass and radius of the earth, respectively, calculate the work carried out on the rocket by the gravitational field of the earth during this part of its flight.
7. A particle undergoes oscillations at the end of a spring of constant $k = 1.2$ N/m. If the maximum amplitude of these oscillations is 10 cm, find how much work is carried out by the spring as the particle goes from:
   (a) Its maximum displacement to one half of this displacement.
   (b) Its equilibrium position to its maximum displacement.
   (c) From a displacement of 3 cm to one of 7 cm.
8. For each of the following potential functions calculate the associated

force $F(x)$. Assume that α, β, V_0 are constant parameters of appropriate dimensions in each case.

(a) $V(x) = \alpha x^2 - \beta x$

(b) $V(x) = \dfrac{\alpha}{x^3}$

(c) $V(x) = V_0 \cos \beta x$

(d) $V(x) = V_0 e^{-\alpha x}$

9. Calculate a potential function $V(x)$ for each of the following force fields. Assume that the quantities α, F_0, and β are fixed constants of suitable dimensions in each case. (a) $F(x) = \alpha x^2 - \beta x^3$; (b) $F(x) = F_0 e^{-\beta x}$; and (c) $F(x) = F_0 \cos \beta x$.

***10.** A spaceship of mass m, traveling through a cloud of galactic dust in outer space, experiences a retarding force $-b\mathbf{v}$, where b is a small positive constant and $\mathbf{v}$ is the velocity of the ship. Set up a coordinate system such that initially at $t = 0$ the spaceship is located at the point x_0, has a velocity v_0, and is traveling along the positive sense of the x-axis.

(a) Write down the equation of motion of the spaceship in this coordinate system, assuming that the viscous force is the only one acting, and show that the velocity $v(t)$ at any time t is

$$v(t) = v_0 e^{-bt/m}$$

(b) Find the position $x(t)$ of the spaceship at any time t.

(c) Calculate the work W carried out on the spaceship by this force from the initial instant $t = 0$ to some later time t when the spaceship is at the position x.

(d) Express your result for W in (c) in terms of the position x at time t and the initial position x_0, and show therefore that this force is *not* conservative.

11. A 100-gram body is thrown upward from the ground with such an initial speed that it achieves a maximum vertical displacement of 2 meters.

(a) What was its kinetic energy on being thrown upward?

(b) What was its kinetic energy after it had risen 1 meter?

(c) What is its kinetic energy upon striking the ground? What happens to this energy?

12. Consider a particle of mass m and total energy E traveling under the action of a conservative force. If $V(x)$ is the potential energy, show that

$$t = \pm \int_{x_0}^{x} \frac{dx}{\left\{\dfrac{2}{m}[E - V(x)]\right\}^{1/2}}$$

where x_0 is the initial position of the particle. (*Hint*: Solve (8-9) for v and integrate.) What is the significance of the ambiguity in sign?

13. A projectile is fired at an angle of elevation α and with a muzzle velocity $\mathbf{v}_0$. Making use of the law of conservation of energy, prove (a) that the magnitude of its velocity at any point in its path is independent of the elevation angle α but varies only with v_0 and (b) that the maximum height h achieved by the projectile is

$$h = \frac{v_0^2 \sin^2 \alpha}{2g}$$

14. Figure 8-22 shows a block of mass m traveling along a *smooth*, horizontal surface with a velocity v_0 when it strikes a spring of constant

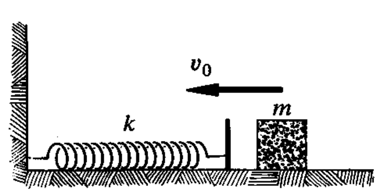


Figure 8-22

k. Calculate the distance d by which the spring is compressed in this process.

15. Consider the Atwood's machine in Figure 8-19 and suppose that frictional effects and the inertia of the wheel may be neglected. Assume that $m_1 = 2$ kg and $m_2 = 3$ kg and that the system is released from rest.

(a) Explain why the sum of the total energies of the two blocks is conserved. (*Hint*: Why does the tension in the supporting string do no work?)

(b) Calculate the speed of the blocks after m_2 has dropped a distance of 2 meters.

16. A bead of mass $m = 10$ grams is free to slide on a *smooth*, circular wire of radius 10 cm and lying in a vertical plane. Assume that it starts to slide down from rest at the top. **(a)** What is its speed when it reaches the lowest point of the wire? **(b)** Calculate its speed when it is at an angular displacement of 90° with respect to the vertical.

17. Consider again the physical situation in Problem 16. Suppose this time, however, that the bead starts out at a velocity of 1.0 m/s at the top of the wire. **(a)** Calculate its speed when it reaches the bottom. **(b)** Calculate its speed when it is at an angular displacement θ with respect to the vertical.

18. A particle of mass m starts out from rest at the top of a smooth, inverted spherical bowl of radius R (see Figure 8-23). Assume that gravity acts vertically downward.

(a) Show that the speed v of the particle when it has slid an angular distance θ down the surface of the bowl is

$$v^2 = 2gR(1 - \cos\theta)$$

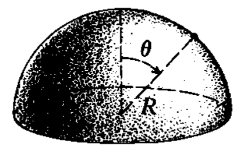


Figure 8-23

(b) Calculate the normal force N exerted by the bowl on the particle as a function of its position θ on the surface of the bowl.

(c) Calculate the angle $\theta = \theta_0$ at which the particle leaves the surface. (*Hint*: The particle will leave the surface at the position for which the normal force N vanishes.)

19. Consider again the situation, shown in Figure 8-23, of a particle of mass m sliding down a smooth, inverted hemispherical bowl of radius R. Assume this time that the particle is initially at the top and starts to slide down with a certain initial speed v_0.

(a) Calculate the speed v of the particle after it has slid an angular displacement θ down the side of the bowl.

(b) Show that the angle θ_0 at which the particle leaves the surface of the bowl is given by the relation

$$\cos\theta_0 = \frac{v_0^2}{3gR} + \frac{2}{3}$$

20. A particle of mass m moves under the combined action of a conservative force, with which is associated a potential energy V and a nonconservative force $\mathbf{F}_v$. Show, by making use of the Newtonian equations of motion, that the time derivative of the sum of the kinetic plus potential energy V of the particle is equal to the rate $\mathbf{v} \cdot \mathbf{F}_v$ at which the force $\mathbf{F}_v$ carries out work on the particle. Describe the circumstances for

which the work carried out by $\mathbf{F}_v$ may vanish, even though this nonconservative force itself does not.

21. A particle starts out at rest at point A at the top of a smooth curved track of vertical height 40 cm (see Figure 8-24). (a) What is its speed at the bottom of the curved track? (b) How far along the adjoining inclined plane, which makes an angle of 30° with the horizontal, will the particle go, assuming that all contacts are smooth?

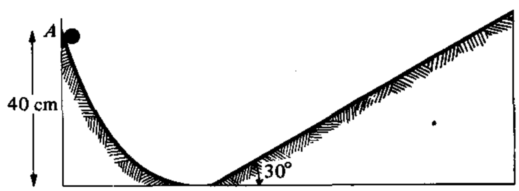


Figure 8-24

22. Suppose in Figure 8-11b a small 50-gram block slides down the track and then travels a distance of 50 cm along a *rough*, horizontal surface before coming to rest. Assuming that the track is smooth and that its height is 30 cm, calculate:
    (a) The velocity of the block when it reaches the bottom of the track.
    (b) The acceleration of the block as it slides along the horizontal surface.
    (c) The coefficient of friction between the block and the horizontal surface.

23. Suppose, in Figure 8-12, that the particle starts out at the top of the inclined plane with a velocity v_0. (a) Assuming all contacts are smooth what is the velocity of the particle at point C in terms of the parameters h, α, v_0, and a? (b) Calculate the normal force on the particle (assuming that its mass is m) at the point A and at the point B.

24. A body of mass m travels in a circular orbit of radius a about a force center characterized by a potential function $V = -k/r^n$, where k and n are fixed, positive constants and r is the radial distance to the particle from the center.
    (a) What is the force on the particle while it is in its circular orbit of radius a?
    (b) Show by applying Newton's second law that the speed v of the particle when in a circular orbit satisfies

$$mv^2 = \frac{nk}{a^n}$$

    (c) Show by use of your result from (b) that the total energy E of the particle is

$$E = \left(1 - \frac{n}{2}\right) V$$

25. A man-made satellite orbits the moon in a very elongated orbit. By use of the conservation-of-energy law prove the following: (a) at the point of closest approach to the center of the moon, its speed assumes a maximum and (b) at the point when it is farthest away from the moon's center, its speed is a minimum.

26. Consider an α particle of mass m that is traveling directly toward a gold nucleus, which *repels* the α particle with the force k/x^2, where x is the separation distance of the α particle from the nucleus and k is a constant. Assume that the velocity of the α particle when very far away from the nucleus is v_0.
    (a) Express the total energy E of the α particle in terms of the given parameters.
    (b) Calculate the distance of closest approach of the α particle to the force center. (*Hint*: What is the velocity of the α particle

at the instant when it is closest to the gold nucleus?)

27. A pendulum of length l has a bob of mass m which is given an initial angular displacement θ_0. If on being released it is given a velocity v_0, show that its speed at a subsequent instant when its angular position is θ is determined by

$$v^2 = v_0^2 + 2gl(\cos\theta - \cos\theta_0)$$

*28. A simple pendulum of length l and mass m is displaced by an angle θ_0 and released at rest. Show, by use of the result in Example 8-10, that the tension in the supporting string is $mg[3\cos\theta - 2\cos\theta_0]$.

29. For the physical situation described in Problem 27 calculate the tension in the string as a function of the angle θ and the parameters l, m, v_0, and θ_0.

30. A simple pendulum of mass m and length l, and hanging vertically at rest, is struck a blow so that it starts to travel horizontally at a velocity $\mathbf{v}_0$. Find its maximum angular displacement.

31. Suppose that a pendulum, consisting of a very light string of length l and a bob of mass m, is originally hanging in equilibrium in a vertical position when it is struck a horizontal blow so that it starts to travel with a velocity $\mathbf{v}_0$. Show that the smallest value that v_0 can have such that the pendulum swings completely around through 180° is

$$v_0 = (5gl)^{1/2}$$

(*Hint*: Why must the tension in the string be zero just as it reaches the top of its path?)

32. For the physical situation described in Problem 30 calculate the tension in the string at the instant immediately following the striking of the bob. Assume the values l = 0.5 meter, m = 400 grams and v_0 = 2.0 m/s.

*33. Consider again the physical situation in Figure 8-14. Calculate the tension P in the string in terms of the angle ϕ and various other parameters shown in the figure.

34. A 10-gram particle is attached to a massless string of length 30 cm and released at rest at an angle of 60° with respect to the vertical, as shown in Figure 8-25. At the point C (defined such that the distance OC is 15 cm and such that the direction OC makes an angle of 45° with respect to the vertical) there is a small peg. Consider the situation after the particle has gone through an angle of 15° so that its subsequent motion may be described by the angle ϕ shown in the figure and by its velocity $\mathbf{v}$.

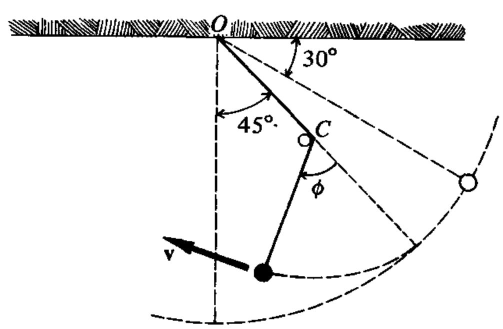


Figure 8-25

(a) What is the value for the total energy of the bob in joules? Take the zero of potential energy at the point of suspension.

(b) What is the velocity of the bob at the instant when the string first makes contact with the peg at point C?

(c) Calculate the maximum value for the angle ϕ and compare this with the prediction Galileo would have made.

35. Repeat all parts of Problem 34, but

this time assume that the bob is given an initial tangential velocity of 1.0 m/s.

36. Consider the motion of a 1-kg body in the potential shown in Figure 8-20 and suppose that its initial kinetic energy when it is at the origin is 3 joules and $V_0 = 1$ joule.
    (a) Calculate the total energy of the particle.
    (b) If it is initially moving to the right, calculate its kinetic energy when it is at a point halfway between x_1 and x_2.
    (c) Is the motion of the particle bounded? Is it periodic?
37. Consider the motion of a body in the potential shown in Figure 8-20, assuming this time that $V_0 =$ 5 joules. Suppose that initially the body is at the point $(x_2 - x_1)/2$, that its kinetic energy is 6 joules at this point, and that it is moving to the right.
    (a) What is its total energy?
    (b) Is the orbit bounded and periodic for these conditions?
    (c) Suppose instead of the above conditions that initially the body is at the point $(x_2 - x_1)/2$ and that it has a kinetic energy of 3 joules. What would be the total energy under these circumstances? Is the orbit bounded now? If so, what are the maximum and minimum values for the position x of the particle?
38. Consider the one-dimensional motion of a body of mass 2 kg with the potential energy shown in Figure 8-26.
    (a) If the body has a kinetic energy of 0.5 joule when it is at the point $x = 1.5$ meter, is the orbit bounded? If not, what is its kinetic energy at the subsequent instant when it is at the point $x = 2.3$ meter?

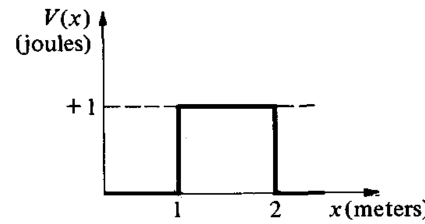


Figure 8-26

    (b) Suppose the body has a kinetic energy of 0.5 joule when it is at the point $x = 0.5$ meter. Can the particle penetrate into the region $x \geqslant 1$ meter?
    (c) Characterize from these facts the nature of the force associated with this potential energy $V(x)$ in Figure 8-26.
39. Consider the one-dimensional motion of a 1-kg body with potential energy $V(x)$ in Figure 8-27.

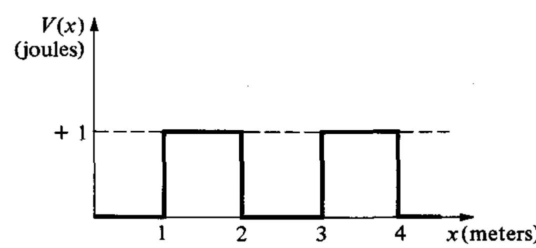


Figure 8-27

    (a) If initially the body is at the point $x = 2.5$ meter and moving to the right with a velocity of 2 m/s, make a plot of the velocity $v(t)$ of the particle as a function of time.
    (b) Suppose instead that when the body is at the point $x =$ 2.5 meter, it is moving to the right with a velocity of 0.5 m/s. Show that in this case the orbit is bounded, and make a plot of the velocity of the body as a function of time for times $t \leqslant$ 5 seconds.

*40. Consider the one-dimensional motion of a body of mass 1 kg in a force field associated with the

potential-energy function

$$V(x) = \frac{1}{x^2} - \frac{2}{x} \quad (x > 0)$$

where all lengths are measured in meters and $V(x)$ is in joules.

(a) Show that the minimum of the potential curve occurs at the point $x = 1$ meter, and make a plot of $V(x)$ as a function of x.

(b) Suppose that the motion of the body in this potential is such that its total energy $E = -0.5$ joule. Draw a horizontal line at this value on your plot and calculate the values for x at which the particle's velocity vanishes.

(c) What is the maximum separation of any two points of the orbit?

41. Consider again the motion of a 1-kg body in the potential of Problem 40. Suppose this time that initially the body is at a very large distance from the origin and moving toward it with a velocity of 2 m/s.

(a) What is the initial kinetic energy of the body?

(b) What is its total energy? What is the kinetic energy when it is at the point $x = 2$ meters? At $x = 1$ meter?

(c) What is the distance of closest approach to the origin that the body can assume? What is its velocity at this point?

†42. A block lying on a horizontal and rough surface is attached to the end of a horizontal spring and oscillates back and forth with ever-decreasing amplitude. If the mass of the block is 200 grams and the coefficient of friction between the block and the surface is 0.1, calculate the rate at which the system is losing energy at an instant when the block is traveling at a speed of 0.5 m/s. At what rate is the interface between the block and the surface being heated up at an instant when its speed is 0.2 m/s?

†43. A block of mass m is attached to the end of a spring and rests in equilibrium on a horizontal surface. Suppose that it is suddenly struck a blow so that it starts to travel at a velocity v_0, and that μ is the coefficient of friction between the two surfaces.

(a) Derive the relation

$$\frac{d}{dt}\left(\frac{1}{2}mv^2 + \frac{1}{2}kx^2\right) = -\mu mgv$$

where v is the instantaneous velocity of the block and x is its position at that instant.

(b) By integrating your result from (a) show that the maximum displacement, A, of the body from equilibrium is given by the solution of the quadratic equation

$$-\frac{1}{2}mv_0^2 + \frac{1}{2}kA^2 = -\mu mgA$$

(c) Find the explicit value for A as predicted by your result from (b). Justify your choice of sign for the root.

9 The law of momentum conservation

This principle is so perfectly general that no particular application of it is possible.

G. POLYA

The momentum of the Universe is constant.

I. NEWTON

9-1 Introduction

In our studies up to this point we have been concerned almost exclusively with the kinematics and dynamics of a single particle. With this chapter we turn from this study in order to begin an analysis of the dynamics of multiparticle physical systems. The specific purpose of this chapter is to consider the special case of an *isolated system.*

As the name implies, an isolated system is one that is not acted upon by forces due to sources *outside of the system itself.* Thus, for an isolated system, the total force acting on a given constituent is due exclusively to the proximity of the other constituents. Strictly speaking, therefore, only the entire universe, considered as an entity in itself, is truly isolated. Nevertheless, there are a variety of interesting physical systems which can also be viewed as being essentially isolated. For example, the system consisting of a proton and an electron bound together to form a hydrogen atom is usually

thought of as being isolated, even though it is acted upon by the external gravitational forces of the earth and of the sun. Similarly, the carts moving on the air track in Figure 4-4 can be treated as the constituents of an isolated system as far as motion along the air track is concerned even though, as we noted in Chapter 4, the gravitational field of the earth and the air track itself exert external forces on the carts.

The main purpose of this chapter is to study a certain conservation law associated with *all* isolated systems. This is the law of *conservation of momentum*, also known as the law of conservation of *linear momentum*. Fully in accordance with this law, experiment shows that the physically realizable motions of the constituents of an isolated system are very severely restricted, and in a way that makes comprehensible several features of the motion of such complex systems. In Section 9-9, for example, this law will be shown to be the important physical principle underlying the operation of jet- and rocket-propulsion systems.

9-2 Momentum conservation of a two-body system

Since it is not possible to obtain a truly isolated system in the laboratory, let us consider in Figure 9-1 two small bodies[1] of respective masses m_1 and m_2 and confined to motion on a smooth, horizontal surface. Suppose that there is a force of interaction of some sort between them, and let us represent this symbolically, as in the figure, by a spring. To the extent that all motion is restricted to the horizontal and smooth surface, the forces of friction and of gravity, which are the only external forces acting on the bodies, play no role. The only force acting on each body therefore is that produced by the other and thus, just as in Section 4-4, they may be thought of as the constituents of an isolated system.

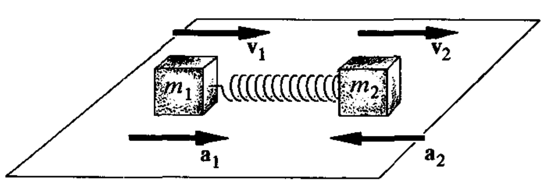


Figure 9-1

Now according to (4-4), regardless of the force between the two bodies—be it electric, magnetic, or by a spring as in Figure 9-1—the accelerations $\mathbf{a}_1$ and $\mathbf{a}_2$ of the two bodies are related by

$$m_1\mathbf{a}_1 + m_2\mathbf{a}_2 = 0$$

[1]Strictly speaking, these should be point particles. However, anticipating the result of Section 9-8, which shows that the arguments here can be extended to other systems, we shall use the terms "particle" and "body" interchangeably.

where m_1 and m_2 are the respective masses of the two bodies. But $\mathbf{a}_1$ and $\mathbf{a}_2$ are the time derivatives of the velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively. Hence, since m_1 and m_2 are the constant masses of the bodies, it follows by use of the definition of a derivative that (4-4) may be expressed in the equivalent form

$$\frac{d}{dt}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2) = 0 \tag{9-1}$$

so that even though in general $\mathbf{v}_1$ and $\mathbf{v}_2$ vary in time, this variation must be such that the quantity $(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)$ has a zero time derivative. In other words, the quantity $(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)$ is a *constant of the motion.*

Therefore if $\mathbf{v}_1^0$ and $\mathbf{v}_2^0$ are the velocities of the two bodies at some initial instant and if $\mathbf{v}_1$ and $\mathbf{v}_2$ are the corresponding velocities at an *arbitrary* subsequent time t, then

$$\boxed{m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}_1^0 + m_2\mathbf{v}_2^0} \tag{9-2}$$

The constancy in time of the quantity $(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)$ is known as the law of the conservation of momentum for this system. In words, this law states:

If m_1 and m_2 are the masses of the two bodies comprising an isolated system, then regardless of the force acting between them and regardless of the nature and shape of the bodies, the quantity $(m_1\mathbf{v}_1 + m_2\mathbf{v}_2)$ is a constant of the motion.

The product of the mass of a body and its velocity, which appears in both (9-1) and (9-2), will play an important role in our studies from now on and will be referred to as the *momentum* or the *linear momentum* of the body. By definition then, the momentum of a body of mass m is

$$\mathbf{p} = m\mathbf{v} \tag{9-3}$$

where $\mathbf{v}$ is its velocity. Thus the momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ of the two bodies in Figure 9-1 are $m_1\mathbf{v}_1$ and $m_2\mathbf{v}_2$, respectively. In terms of $\mathbf{p}_1$ and $\mathbf{p}_2$, the content of (9-1) or, equivalently, (9-2), is that the quantity $(\mathbf{p}_1 + \mathbf{p}_2)$, which is the total momentum of the isolated two-body system, is a constant of the motion. If $\mathbf{p}_1^0$ and $\mathbf{p}_2^0$ are the initial momenta of the two bodies comprising the system, and $\mathbf{p}_1$ and $\mathbf{p}_2$ the corresponding momenta at an arbitrary subsequent instant, then

$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_1^0 + \mathbf{p}_2^0 \tag{9-4}$$

In other words, the total momentum of the two bodies has at all times the same value that it has at the initial instant.

It deserves to be emphasized that the law of momentum conservation as expressed in (9-4) is very general; no experimental violation of it has ever been found under any circumstance for any truly isolated system. It is valid not only for macroscopic bodies but also for microscopic systems, for

some of which Newton's laws themselves must be replaced by those of quantum mechanics. In this respect the law of momentum conservation for isolated systems differs in an important way from that of the law of conservation of energy. The energy of a system is conserved only if the forces acting on the constituent members of the isolated system are conservative, whereas (9-2) is invariably correct for such systems. Moreover, the law of momentum conservation in the form (9-4) is applicable also for relativistic velocities, when the speed of one member of the system is close to that of light. In this case, however, it is necessary to modify the definition of momentum, (9-3), in an appropriate way. We shall consider the matter further in Section 9-10.

Example 9-1 Suppose, in Figure 9-1, that $m_1 = 2$ kg and $m_2 = 1$ kg, and that initially the two bodies are in equilibrium and at rest. If m_2 is suddenly struck a blow and starts to travel to the right with a velocity $v_0 = 2$ m/s, what is the velocity of m_1 at a subsequent instant when the velocity of m_2 is:
(a) Zero?
(b) 0.5 m/s, directed to the left?

Solution Let v_1, v_2, and v_2^0 be the components of the various velocities in Figure 9-1 along the direction of motion. Since $v_1^0 = 0$ and $v_2^0 = 2$ m/s, the conservation-of-momentum relation, (9-2), becomes

$$m_1 v_1 + m_2 v_2 = m_2 v_2^0 = 2 \text{ kg-m/s}$$

Dividing through by m_1 and using the known values for m_1 and m_2, we obtain

$$v_1 + \frac{1}{2} v_2 = 1 \text{ m/s}$$

(a) Substituting into this the value $v_2 = 0$ we find that

$$v_1 = 1 \text{ m/s}$$

and thus at the instant when m_2 is at rest, m_1 has a velocity of 1 m/s, directed to the right.

(b) For this case, $v_2 = -0.5$ m/s, and substituting this into the above relation, we find that

$$v_1 = 1.25 \text{ m/s}$$

This means that the two bodies are on a collision course at this instant. Note that the force between the two bodies plays no role in this calculation.

Example 9-2 Consider, in Figure 9-2, a man with a pistol standing on a horizontal sheet of ice. If the total mass of the man, the pistol, and its bullets is 80 kg and the mass of a bullet is 20 grams, and if after firing a shot as in Figure 9-2b he recoils with a velocity of $v_2 = 15$ cm/s, directed to the right, what is the muzzle velocity of the bullet? Neglect friction.

Solution Let us consider the man and the pistol as one constituent and the bullet as the second constituent of an isolated two-body system. According to the

conservation-of-momentum relation, since the total momentum of the system before firing is zero we may write

$$m_1 v_1 + m_2 v_2 = 0$$

where v_1 is the component of the velocity of the bullet along an axis assumed to be directed to the right in the figure, v_2 is the corresponding component of the recoil velocity of the man, $m_1 = 20$ grams is the mass of the bullet, and $m_2 = 80$ kg ($\simeq 80\text{ kg} - 0.02\text{ kg}$) is the total mass of the man and the gun. Solving for v_1 and, substituting the known values for the remaining parameters, we find that

$$\begin{aligned} v_1 &= -v_2 \frac{m_2}{m_1} = -(0.15\text{ m/s}) \left(\frac{80\text{ kg}}{0.02\text{ kg}}\right) \\ &= -600\text{ m/s} \end{aligned}$$

where the minus sign reflects the fact that the direction of motion of the bullet is opposite to that of the motion of the man.

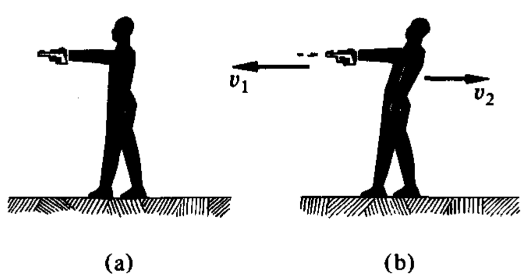


Figure 9-2

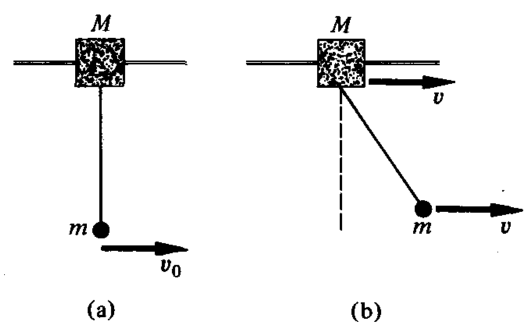


Figure 9-3

Example 9-3 Consider, in Figure 9-3, a simple pendulum, whose bob has a mass m and is suspended from a block of mass M, which in turn is free to slide along a smooth, horizontal rail. If the bob is suddenly struck a blow so that it starts to travel in a horizontal direction with a velocity v_0, calculate its velocity at a subsequent instant when it has risen to its maximum height.

Solution Since the force of gravity is an external force acting on this two-body system, strictly speaking, it is not isolated. However, since this external force acts only along the vertical direction, it follows that as far as motion along the horizontal is concerned, the system is indeed isolated and the component of the total momentum along the horizontal direction must be conserved.

Initially, the total momentum of this system is mv_0 since the block has no momentum at this instant. Subsequently, when the bob has reached its maximum elevation, it will have no motion relative to its supporting block and thus both will be traveling along the horizontal direction at some unknown velocity v (see Figure 9-3b). Accordingly, since the component of momentum along the horizontal direction is conserved, it follows that

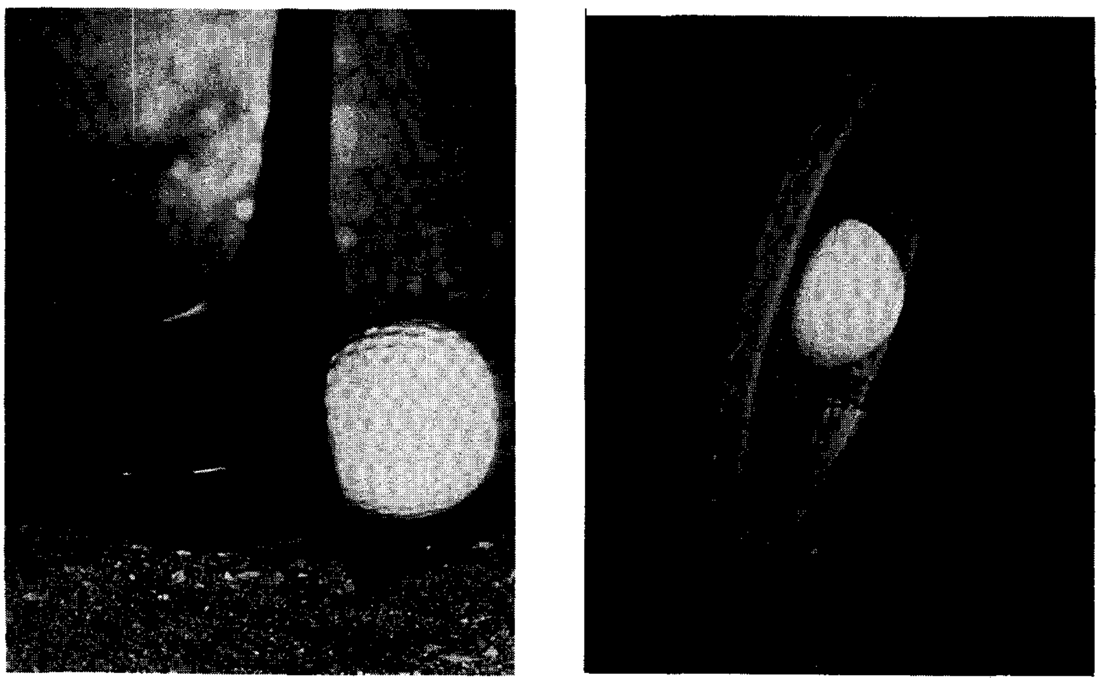
$$mv_0 = mv + Mv$$

and this leads to

$$v = v_0 \frac{m}{m + M}$$

9-3 Impulsive forces—qualitative

Physical situations often arise for which the forces that the constituents of a system exert on each other last for only a very short time. However, during the interval while they last, their strength is so great that they produce an easily observable effect. We call forces of this type *impulsive*. The force that a golf ball or a tennis ball experience at the instant they are struck, as in Figure 9-4, is one example of an impulsive force. The force that the fragments of an exploding shell exert on each other during the instant of the explosion is a second example of such a force. A third is illustrated in Figure 9-5, which shows the collision of an antiproton ($\bar{p}$) with a proton (p), out of which emerge a cascade hyperon pair. Here, the force that the $\bar{p}$ and the p exert on each other while they are momentarily in contact is the impulsive force.



(a) (b)

Figure 9-4 *Flash photograph of a golf club and ball (a) and of a tennis racket and ball (b) at the instant of collision. Note the extent of the deformations, which are a measure of the strength of the impulsive forces involved. (Courtesy Harold E. Edgerton.)*

To obtain some feeling for the orders of magnitude of the physical quantities involved, let us consider, by way of illustration, the collision of two atoms. Suppose, as shown schematically in Figure 9-6, that an atom of mass m approaches with a speed v_0 an identical atom initially at rest and that after the collision the second atom moves to the right with the same velocity $\mathbf{v}_0$. Studies in atomic physics have shown that the force between two atoms vanishes for separation distances larger than about 5×10^{-10} meter, whereas for shorter distances this variation is roughly that shown in Figure 8-15b. Accordingly,

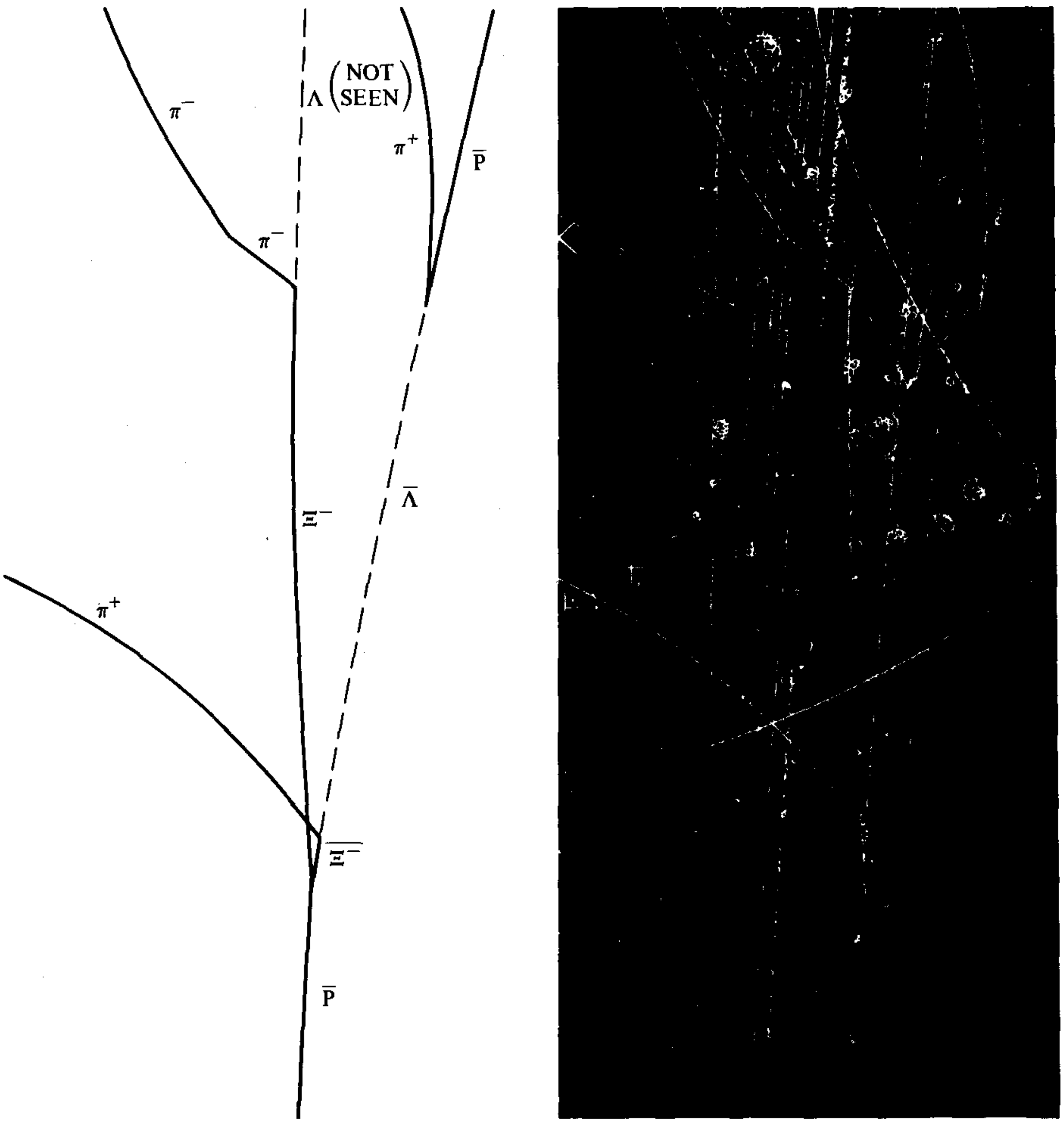


***Figure 9-5** A hydrogen bubble-chamber photograph and explanatory sketch showing the collision of an antiproton and a proton resulting in the production of a cascade hyperon pair. (See Table 1-2.) (Courtesy Brookhaven National Laboratory.)*

as the atom in Figure 9-6a approaches to within a distance of about 5×10^{-10} meter of the target atom, it becomes aware of the interatomic repulsive force and starts to slow down. At the same time the target atom experiences the same force but oriented in the opposite direction, and thus it starts to move to the right. Finally, as shown in Figure 9-6b, the target atom travels to the right with the velocity $\mathbf{v}_0$ and the incident atom comes to rest.

These changes in the velocities of the two atoms do not take place instantaneously, of course, but occur over a certain finite time interval; let us call it Δt_0. As a first estimate for Δt_0 we may take it to be the ratio of the range of the interatomic force and the incident velocity v_0. Assuming a speed

Figure 9-6

of 5×10^3 m/s, we find that

$$\Delta t_0 \cong \frac{\text{force range}}{v_0} = \frac{5 \times 10^{-10}\text{ m}}{5 \times 10^3\text{ m/s}}$$
$$= 10^{-13}\text{ s}$$

This means that the change in the velocity of the target atom from its initial value of zero to the final value of 5×10^3 m/s takes place in a time interval of the order of 10^{-13} second, and corresponds to an acceleration of magnitude $v_0/\Delta t_0 \cong 5 \times 10^{16}\text{ m/s}^2$. This is sixteen orders of magnitude larger than the acceleration of gravity! Thus, to produce the effect in Figure 9-6, the force between the atoms, which acts only for about 10^{-13} second, must be very large indeed.

It is these two features of a force—namely, of its being very strong and of its having nonzero values only for a very short time interval—that characterize *impulsive* forces in general.

9-4 Impulse

Let us now discuss impulsive forces more quantitatively. Consider a particle of mass m moving initially at some constant velocity $\mathbf{v}_i$ when at a certain instant t_i it begins to experience an impulsive force $\mathbf{F}_I(t)$. Figure 9-7a shows a typical variation in time of a component $F_I(t)$ of such a force. We assume that the impulsive force is zero for times t less than t_i, that it rises very quickly to a large value, and then decreases very rapidly, so that at a certain later time t_f it is again zero. Because of the impulsive nature of $\mathbf{F}_I$, the time interval $\Delta t_0 = t_f - t_i$ during which the force acts is assumed to be impercepti-

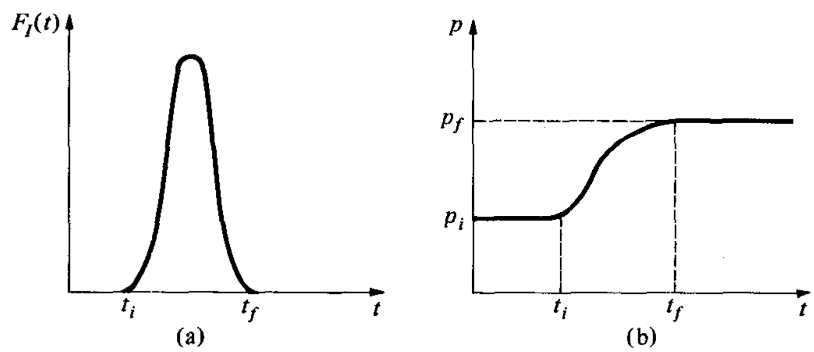


Figure 9-7

bly small. Figure 9-7b shows a plot of the corresponding component of the momentum p (or the velocity v) of the particle as a function of time. For times $t < t_i$ the momentum of the particle has the value $\mathbf{p}_i$, whereas for times $t > t_f$, when the impulsive force no longer acts, the particle has a certain final momentum $\mathbf{p}_f$. Between these times while the impulsive force $\mathbf{F}_I(t)$ acts, the particle's momentum will vary, in a generally very complicated manner, depending on the details of the impulsive force itself.

To obtain a quantitative measure of an impulsive force, we start with Newton's law of motion, which, by use of (9-3), may be expressed in the form

$$\mathbf{F}_I = \frac{d\mathbf{p}}{dt} \tag{9-5}$$

since the mass m of the body is constant. Integrating over the time interval Δt_0, during which the impulsive force acts, we have

$$\int_{t_i}^{t_f} \mathbf{F}_I\, dt = \int_{t_i}^{t_f} \frac{d\mathbf{p}}{dt}\, dt = \int_{\mathbf{p}_i}^{\mathbf{p}_f} d\mathbf{p}$$
$$= \mathbf{p}_f - \mathbf{p}_i$$

where the second equality follows by changing the variable of integration from t to $\mathbf{p}$, and where $\mathbf{p}_i = m\mathbf{v}_i$ is the particle's momentum just before $\mathbf{F}_I$ begins to act and $\mathbf{p}_f = m\mathbf{v}_f$ is the corresponding momentum afterwards. The integral on the left, which represents the area under the $F_I(t)$ curve, is known as the *impulse* $\mathbf{J}$. It is defined by

$$\mathbf{J} = \int_{t_i}^{t_f} \mathbf{F}_I(t)\, dt \tag{9-6}$$

and its units are the same as those of momentum: that is, the N-s = kg-m/s. The substitution of $\mathbf{J}$ as given in (9-6) into the above formula leads to

$$\mathbf{J} = \mathbf{p}_f - \mathbf{p}_i \tag{9-7}$$

which, by the use of (9-3), may be expressed alternatively as

$$\mathbf{J} = m\mathbf{v}_f - m\mathbf{v}_i \tag{9-8}$$

As a general rule it is difficult to measure the impulse $\mathbf{J}$ associated with an impulsive force directly. However, according to (9-7) the impulse $\mathbf{J}$ imparted to a body is equal *to the change in momentum of the body.* Therefore, since all three of the quantities, m, $\mathbf{v}_f$, and $\mathbf{v}_i$, are each separately measureable, it follows from (9-8) that the impulse $\mathbf{J}$ associated with any impulsive force can be ascertained indirectly by observing its effect in changing the momentum of the body on which it acts.

Generally speaking, impulsive forces are sufficiently strong so that during the short time interval during which they act, all other forces acting on the particle may be neglected. Hence (9-7) is applicable, for example, even if the particle were acted upon by the gravitational field of the earth in addition to $\mathbf{F}_I(t)$. In these cases, however, it is important to keep in mind that $\mathbf{p}_i$ and $\mathbf{p}_f$

refer respectively to the momenta *just before* and *just after* the time interval that the impulsive force acts. Outside of this time interval nonimpulsive forces cannot be neglected.

Example 9-4 A 100-gram hockey puck is at rest on a frozen lake when it is struck a blow which causes it to travel along the ice at a speed on 30 m/s.

(a) What impulse has been imparted to the hockey puck?

(b) If the length of time that the hockey stick and the puck are in contact is 10^{-3} second, what is the approximate magnitude of the impulsive force?

Solution

(a) According to (9-8), the impulse J imparted to the hockey puck is equal to the change in its momentum. Hence, since $v_i = 0$ in this case,

$$J = p_f - p_i = mv_f - mv_i = 0.1\text{ kg} \times 30\text{ m/s} - 0$$
$$= 3.0\text{ kg-m/s}$$

(b) An estimate for the force acting on the puck may be obtained by dividing the impulse J by the length of time Δt_0 during which the force acts. Since this time has been given to be 10^{-3} second, it follows that

$$F_I = \frac{J}{\Delta t_0} = \frac{3.0\text{ kg-m/s}}{10^{-3}\text{ s}} = 3.0 \times 10^3\text{ N}$$

Example 9-5 A ball of mass m is dropped from a height h above the ground (see Figure 9-8) and rebounds to a height $h/4$.

(a) What was its velocity $\mathbf{v}_i$ just prior to striking the ground?

(b) What is its velocity $\mathbf{v}_f$ just after it bounces?

(c) What impulse $\mathbf{J}$ has been imparted to the ball?

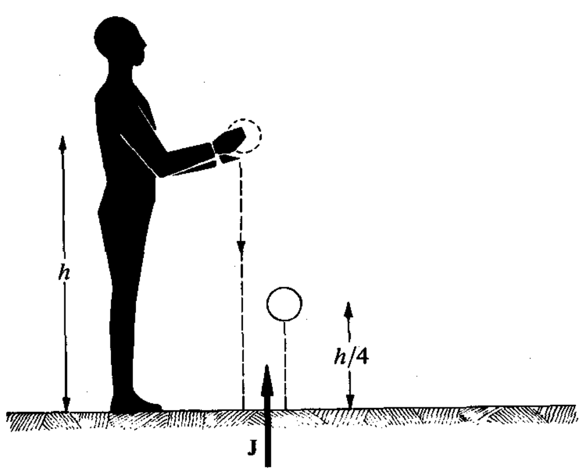


Figure 9-8

Solution

(a) Making use of (2-18), we find that, since $v_0 = 0$ and $x = 0$ in this case,

$$\mathbf{v}_i = -(2gh)^{1/2}\mathbf{j}$$

where $\mathbf{j}$ is a unit vector directed vertically upward.

(b) Since, after rebounding, the ball rises to a height $h/4$, its velocity $\mathbf{v}_f$ is found in a similar way to be

$$\mathbf{v}_f = \left(\frac{2gh}{4}\right)^{1/2}\mathbf{j} = \frac{1}{2}(2gh)^{1/2}\mathbf{j}$$

(c) Substituting these values for the velocity into (9-8), we find that

$$\begin{aligned}\mathbf{J} &= m(\mathbf{v}_f - \mathbf{v}_i)\\ &= \frac{3}{2}m(2gh)^{1/2}\mathbf{j}\end{aligned}$$

Example 9-6 A proton ($m = 1.67 \times 10^{-27}$ kg) initially traveling at a speed of 10^6 m/s collides with an alpha particle ($M = 6.67 \times 10^{-27}$ kg) initially at rest. If after the collision the alpha particle is traveling at a velocity $v_\alpha = 2.5 \times 10^4$ m/s along the original direction of motion of the proton, calculate:

(a) The impulse $\mathbf{J}_\alpha$ imparted to the alpha particle.
(b) The impulse $\mathbf{J}_p$ imparted to the proton in this collision.
(c) The final velocity $\mathbf{v}_f$ of the proton.

Solution

(a) Assuming that all motion takes place along a single direction and that we deal only with components along this direction, we find by use of (9-8) that since initially the alpha particle is at rest, then

$$\begin{aligned}J_\alpha = Mv_\alpha - 0 &= 6.67 \times 10^{-27}\ \text{kg} \times 2.5 \times 10^4\ \text{m/s}\\ &= 1.67 \times 10^{-22}\ \text{kg-m/s}\end{aligned}$$

(b) According to Newton's third law, the force that the α particle exerts on the proton during the collision must be equal and opposite to that which the proton exerts on the α particle. It follows that the impulse to the proton $\mathbf{J}_p$ must be equal and opposite to that imparted to the α particle. Hence

$$J_p = -1.67 \times 10^{-22}\ \text{kg-m/s}$$

where the minus sign shows that the direction of the impulse to the proton is opposite to that imparted to the α particle.

(c) Applying (9-8) to the proton, we have

$$J_p = m(v_f - v_i)$$

where v_f is the sought-for final velocity of the proton and $v_i = 10^6$ m/s is the initial value. Solving for v_f and inserting the known values of the other parameters, we obtain

$$v_f = v_i + \frac{J_p}{m} = 10^6\ \text{m/s} - \frac{1.67 \times 10^{-22}\ \text{kg-m/s}}{1.67 \times 10^{-27}\ \text{kg}} = 9.0 \times 10^5\ \text{m/s}$$

9-5 Two-body collisions

Consider an isolated two-body system. We say that the bodies undergo a *collision*, provided that circumstances are such that the force between them

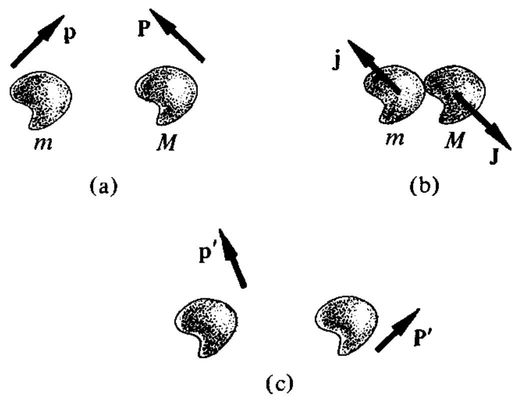


Figure 9-9

is of an impulsive nature. In Figure 9-4, for example, at the instant shown, the golf ball and the club undergo a collision since the force between them is very large and lasts only for an imperceptibly short interval of time. The purpose of this section is to consider collision processes in light of the ideas of momentum conservation.

Consider, in Figure 9-9a, two bodies of masses m and M and momenta $\mathbf{p}$ and $\mathbf{P}$, traveling at the respective velocities $\mathbf{v}$ $(= \mathbf{p}/m)$ and $\mathbf{V}$ $(= \mathbf{P}/M)$ prior to the collision. Assuming that they collide, let $\mathbf{j}$ and $\mathbf{J}$ be the respective impulses imparted to the two bodies and, as shown in Figure 9-9c, let their respective momenta after the collision be $\mathbf{p}'$ and $\mathbf{P}'$. According to (9-7), the impulses $\mathbf{j}$ and $\mathbf{J}$ imparted to the two bodies during the collision are related to their momenta by

$$\begin{aligned} \mathbf{j} &= \mathbf{p}' - \mathbf{p} \\ \mathbf{J} &= \mathbf{P}' - \mathbf{P} \end{aligned} \tag{9-9}$$

Further, according to Newton's third law, the forces which the two bodies exert on each other while they are colliding must be equal and opposite. Therefore, as shown in Figure 9-9b, the impulses $\mathbf{j}$ and $\mathbf{J}$ on the two bodies must also be equal and opposite, so that we have

$$\mathbf{j} = -\mathbf{J}$$

Substituting (9-9) into this relation we find, after rearranging terms, that

$$\mathbf{p} + \mathbf{P} = \mathbf{p}' + \mathbf{P}' \tag{9-10}$$

which is formally similar to (9-4) and states that the total momentum of the system $(\mathbf{p} + \mathbf{P})$ before the collision is the same as the total momentum $(\mathbf{p}' + \mathbf{P}')$ afterward. The conservation-of-momentum relation in (9-4) is more general, however, for it states that momentum is conserved during the collision as well.

In terms of the velocities $\mathbf{v}$ and $\mathbf{V}$ before the collision and $\mathbf{v}'$ and $\mathbf{V}'$ afterward, (9-10) may be expressed equivalently as

$$m\mathbf{v} + M\mathbf{V} = m\mathbf{v}' + M\mathbf{V}' \tag{9-11}$$

where $\mathbf{v} = \mathbf{p}/m$, $\mathbf{V} = \mathbf{P}/M$, and so forth. Again, this is formally the same as (9-2), but with a slightly different interpretation for the velocities.

If the sum of the kinetic energies of the colliding bodies in the system is the same before the collision as it is afterward, the collision is said to be *elastic*. Collisions that involve a loss of kinetic energy are *inelastic*. Hence, if the collision in Figure 9-9 is elastic, then

$$\frac{1}{2}m\mathbf{v}^2 + \frac{1}{2}M\mathbf{V}^2 = \frac{1}{2}m\mathbf{v}'^2 + \frac{1}{2}M\mathbf{V}'^2 \tag{9-12}$$

For elastic collisions, then, (9-12) and the three component relations implied by (9-11) yield four relations among the twelve components of the velocity vectors of the two particles. The specification of any eight of these yields unique values for the remaining four.

It must be emphasized that, at least for isolated systems, regardless of the nature of the forces between the bodies or their shapes, the law of momentum conservation is *always valid*. On the other hand, the same statement *cannot* be made for the total kinetic energy of the system. As a general rule, collisions involving macroscopic bodies such as baseballs, stones, or hammers are inelastic, since they are usually accompanied by energy losses in the form of heat and sound. For these cases, the initial kinetic energy always exceeds the final value by the amount dissipated as heat or sound. On the other hand, collisions at the microscopic level can be either elastic or inelastic. For low enough energies, collisions involving, say, electrons or protons or the nuclei of atoms are normally elastic, but as the kinetic energies of the colliding particles are raised, inelastic processes also begin to take place. Thus, above a certain threshold energy the collision of two protons is inelastic since it may be accompanied by the production of a neutral pion:

$$p + p \rightarrow p + p + \pi^\circ$$

Similarly, collisions involving atoms and molecules will be elastic if the internal energy states of these particles are not excited in the process. As a general rule, the lower the kinetic energies of the atoms and molecules involved in a collision the smaller is the probability of an inelastic collision taking place.

In working collision problems, it must be borne in mind that unless the process is known *a priori* to be elastic, or approximately so, (9-12) cannot be used. By contrast, the law of momentum conservation applies to both elastic and inelastic collisions.

9-6 Applications

In this section the general utility of the law of momentum conservation to collision problems will be illustrated by applying it to a variety of physical

situations. Unless a statement is made to the contrary, it is necessary in each case to assume that the collision is inelastic. In some cases the energy lost in the collision can be calculated by use of the momentum conservation principle.

Example 9-7 Suppose, in Figure 9-6a, that a proton of mass m travels at a velocity $\mathbf{v}_0$ and collides with a second identical proton, initially at rest. If, after the collision, the original proton is traveling at a speed $v_0/4$ along the original direction of motion, calculate the final velocity $\mathbf{V}'$ of the target proton. How much energy is lost in the collision?

Solution In the notation of (9-11), we are given the values $m = M$, $v = 0$, $V = v_0$, and $v' = v_0/4$. Substituting these values, (9-11) becomes

$$0 + mv_0 = m\left(\frac{1}{4}v_0\right) + mV'$$

and solving for V' we find that

$$V' = \frac{3}{4}v_0$$

The energy loss ΔE in this collision is the difference between the initial and final kinetic energies; thus

$$\Delta E = \frac{1}{2}mv_0^2 - \frac{m}{2}\left[\left(\frac{1}{4}v_0\right)^2 + \left(\frac{3}{4}v_0\right)^2\right]$$
$$= \frac{3}{16}mv_0^2$$

Thus there is a definite energy loss of the system, and the collision is inelastic.

Example 9-8 A freight car of mass 3.0×10^4 kg rolls on a siding at 10 km/hr when it collides with a second freight car of mass 6.0×10^4 kg at rest. Calculate the velocity of the freight cars after the collision, assuming that they are coupled together as a result.

Solution In the notation of (9-11), the parameter values are $m = 3.0 \times 10^4$ kg, $M = 6.0 \times 10^4$ kg, $v = 10$ km/hr, $V = 0$. Calling the common final velocity of the two freight cars v', we obtain

$$3.0 \times 10^4\ \text{kg} \times 10\ \text{km/hr} + 0 = v'(3.0 \times 10^4\ \text{kg} + 6.0 \times 10^4\ \text{kg})$$

and this simplifies to

$$v' = 3.3\ \text{km/hr}$$

It is left as an exercise to confirm that this collision is also inelastic.

Example 9-9 A cannonball at rest explodes into two fragments, one of which is four times more massive than the other. If, immediately after the explosion, the lighter fragment is observed to be traveling due north with a speed of 300 m/s, what is the magnitude and direction of the velocity of the other fragment at that instant?

Solution Before the explosion, the cannonball can be thought of as consisting of two fragments of respective masses $M_0/5$ and $4M_0/5$, where M_0 is the total mass of the cannonball. Using the given data, we make the identification: $m = M_0/5$, $M = 4M_0/5$, $v = V = 0$, and $v' = 300$ m/s. Substituting into (9-11), we find that

$$0 + 0 = \frac{1}{5} M_0 \times 300 \text{ m/s} + \frac{4}{5} M_0 V'$$

Thus the velocity V' of the heavier fragment is

$$V' = -75 \text{ m/s}$$

where the minus sign signifies that this fragment is traveling due south—that is, in the direction opposite to that of the lighter fragment.

Note that, in this case, the energy of the two bodies after the "collision" is *greater* than the initial energy. This excess energy is added to the system by the energies generated by the exploding chemicals. Note also that (9-11) was used even though the bodies involved in this collision were *not* isolated. The justification for this is based on the fact that the force of gravity on the fragments is negligible compared to the impulsive force of the explosion. Thus (9-11) is applicable, but only just before and just after the explosion.

Example 9-10 A 15-gram bullet is traveling horizontally at a speed of 200 m/s when it strikes the bob of a pendulum of mass 1 kg (see Figure 9-10). If the bob is observed to be traveling at a speed of 2 m/s immediately after the collision, calculate the speed with which the bullet emerges from the opposite side of the bob. Neglect any motion of the bob during the transit time of the bullet through it.

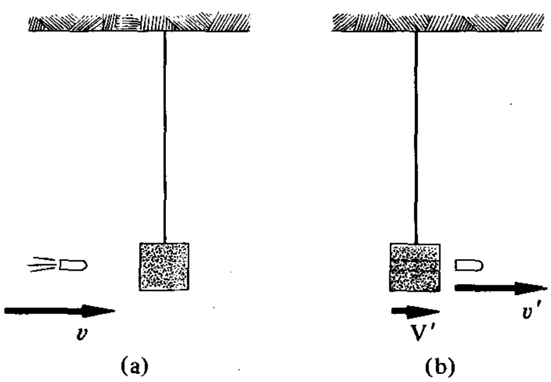


Figure 9-10

Solution The parameter values here are $m = 15$ grams, $M = 1$ kg, $v = 200$ m/s, $V = 0$, and $V' = 2$ m/s. Substituting these values into (9-11), we obtain

$$(0.015 \text{ kg} \times 200 \text{ m/s}) + 0 = (1 \text{ kg} \times 2 \text{ m/s}) + (0.015 \text{ kg} \times v')$$

and this leads to

$$v' = 67 \text{ m/s}$$

Example 9-11 A particle of mass 5 grams is traveling at speed of 10 cm/s when it collides with a second particle of mass 10 grams, originally at rest. If, after the

collision, the 5-gram particle has a speed of 6 cm/s in a direction making an angle of 45° with respect to the incident direction (see Figure 9-11), what is the velocity of the 10-gram particle after the collision and how much energy is lost?

Solution Let us take a coordinate system in the plane of scattering, as shown in Figure 9-11b, with the x-axis along the direction of the incident velocity. If θ is the angle the velocity of the 10-gram particle makes with the direction of the x-axis, and if we assume that all masses are measured in grams and all velocities in centimeters per second, then since the momentum along the x-direction must be conserved, it follows that

$$0 + 5 \times 10 = 5 \times 6 \times \cos 45° + 10v \cos \theta$$

The corresponding relation along the y-direction is

$$0 = 5 \times 6 \times \sin 45° - 10v \sin \theta$$

Solving for $v \cos \theta$ and $v \sin \theta$, we find

$$v \cos \theta = 2.9$$
$$v \sin \theta = 2.1$$

and dividing the first of these into the second, the unknown velocity v cancels, and we find that

$$\tan \theta = \frac{2.1}{2.9} \cong 0.72$$

and thus $\theta = 36°$. Finally, the substitution of this value for θ leads to

$$v = 3.6 \text{ cm/s}$$

To calculate the energy loss, we note that the 5-gram particle has all of the initial kinetic energy. Thus the initial kinetic energy is

$$\frac{1}{2} mv^2 = \frac{1}{2} \times 5.0 \times 10^{-3} \text{ kg} \times (0.1 \text{ m/s})^2 = 2.5 \times 10^{-5} \text{ J}$$

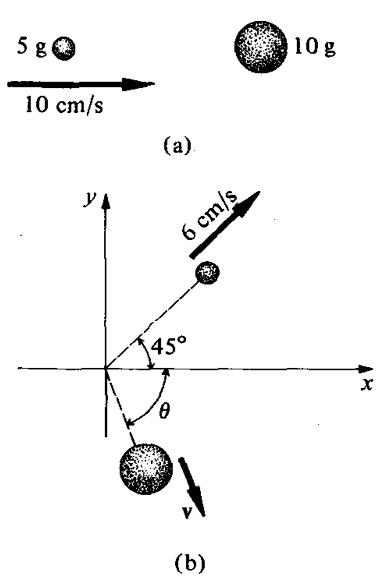


Figure 9-11

Afterward, the 5-gram particle has a velocity of 0.06 m/s and the 10-gram particle has a velocity of 0.036 m/s. Hence the energy after the collision is

$$\frac{1}{2}mv'^2+\frac{1}{2}MV'^2=\frac{1}{2}\times 5\times 10^{-3}\,\text{kg}\times(0.06\,\text{m/s})^2+\frac{1}{2}\times 10^{-2}\,\text{kg}\times(0.036\,\text{m/s})^2$$
$$=1.6\times 10^{-5}\,\text{J}$$

This implies that

$$2.5\times 10^{-5}\,\text{J}-1.6\times 10^{-5}\,\text{J}=0.9\times 10^{-5}\,\text{J}$$

of energy is lost in this process, and thus the collision is inelastic.

9-7 Center-of-mass motion

Consider, in Figure 9-12, an isolated system consisting of two particles of masses m_1 and m_2 with positions, relative to an origin O, of $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. Since the system is isolated, it follows that its total momentum $\mathbf{P}$, defined by

$$\mathbf{P}=\mathbf{p}_1+\mathbf{p}_2=m_1\mathbf{v}_1+m_2\mathbf{v}_2 \tag{9-13}$$

where $\mathbf{v}_1=d\mathbf{r}_1/dt$ and similarly for $\mathbf{v}_2$, is conserved. The purpose of this section is to view this result from a new perspective.

To this end, let us define, for this system, a certain position vector $\mathbf{R}_c$, called the *center of mass*, by

$$\mathbf{R}_c=\frac{m_1\mathbf{r}_1+m_2\mathbf{r}_2}{m_1+m_2} \tag{9-14}$$

The physical significance of $\mathbf{R}_c$ is illustrated in Figure 9-13, which shows that $\mathbf{R}_c$ is a vector drawn from the origin O to a point on the line joining the two particles and dividing it in the ratio m_1 to m_2. In other words, $\mathbf{R}_c$ divides the line joining the two particles into two segments of respective lengths $m_2|\mathbf{r}_1-\mathbf{r}_2|/(m_1+m_2)$ and $m_1|\mathbf{r}_1-\mathbf{r}_2|/(m_1+m_2)$. Generally speaking, the center of mass is closer to the more massive of the two particles.

Let us now consider the physical significance of the center of mass of an isolated two-body system in light of the law of momentum conservation. To this end, let us multiply both sides of (9-14) by the total mass (m_1+m_2),

$$(m_1+m_2)\mathbf{R}_c=m_1\mathbf{r}_1+m_2\mathbf{r}_2$$

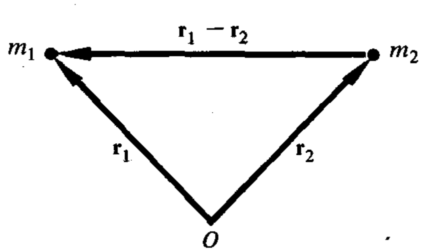


Figure 9-12

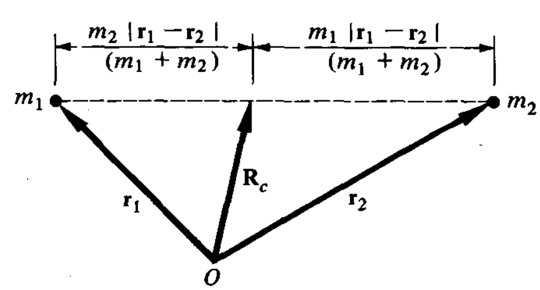


Figure 9-13

Taking the time derivative, we obtain

$$(m_1 + m_2)\mathbf{V}_c = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 \tag{9-15}$$

where $\mathbf{V}_c (\equiv d\mathbf{R}_c/dt)$ is the velocity *of the center of mass.* As the two particles accelerate under their mutual interactions, their center of mass will travel at this velocity $\mathbf{V}_c$. However, since the total momentum $\mathbf{P}$ is a constant of the motion, it follows, by comparing (9-15) and (9-13), that the velocity of the center of mass $\mathbf{V}_c$ is also constant in time. Thus we have established that

---

The center of mass of an isolated two-body system travels with uniform velocity in a straight line.

In other words, the motion of the center of mass of a two-body system is precisely that of a single particle as expressed by Newton's first law. Just as an isolated particle will continue to move with uniform velocity in a straight line, in a similar way an isolated two-body system will move in such a way *that its center of mass moves with uniform velocity in a straight line.*

It should be noted that the center of mass of an isolated two-particle system is a geometric point that lies along the line joining the two particles. In general, no matter of any type need be at the point itself. Thus $\mathbf{R}_c$ is a fictitious point in space which lies somewhere between the particles; and even though the particles themselves may move in a very complex way, the motion of the point described by $\mathbf{R}_c$ is extremely simple.

Example 9-12 Find the center of mass of two bodies of respective masses 3 kg and 2 kg, located as shown in Figure 9-14.

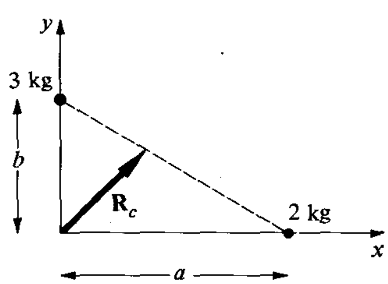


Figure 9-14

Solution Using the subscript "1" for the 3-kg particle and the subscript "2" for the 2-kg particle, we may express the position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ as

$$\mathbf{r}_1 = b\mathbf{j} \qquad \mathbf{r}_2 = a\mathbf{i}$$

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors along the appropriate axes. Substituting these values into (9-14), we find

$$\mathbf{R}_c = \frac{3\text{ kg} \times b\mathbf{j} + 2\text{ kg} \times a\mathbf{i}}{2\text{ kg} + 3\text{ kg}} = \frac{3}{5}b\mathbf{j} + \frac{2}{5}a\mathbf{i}$$

It is left as an exercise to confirm the fact that this vector $\mathbf{R}_c$ lies along the line joining the two particles and divides it in the ratio of 3 to 2, as it must.

Example 9-13 Calculate the center of mass of the earth and the sun, assuming them to be point particles.

Solution According to Figure 9-13 the center of mass divides the line joining the earth and the sun in the ratio

$$\frac{M_E}{M_S} = \frac{6.0 \times 10^{24}\ \text{kg}}{2.0 \times 10^{30}\ \text{kg}} = 3.0 \times 10^{-6}$$

Thus, since the distance d from the center of the sun to the center of the earth is 1.5×10^8 km, it follows that the distance of the center of mass from the sun is

$$3.0 \times 10^{-6} \times 1.5 \times 10^8\ \text{km} = 450\ \text{km}$$

Since the sun's radius is of the order of 7×10^5 km, the center of mass of the earth-sun system is well within the interior of the sun!

Example 9-14 Two bodies of respective masses $m_1 = 1$ kg and $m_2 = 2$ kg are connected by a spring and are originally in equilibrium on a horizontal and smooth surface at a separation distance of 0.5 meter, as shown in Figure 9-15. Assume that the motion of the bodies is confined to the x-axis.

(a) What is the initial location of the center of mass?

(b) If m_1 is suddenly struck a blow so that it starts to move to the right with a speed of 2 m/s, what is the initial velocity of the center of mass at this instant?

(c) What is the position of the center of mass at any subsequent time t?

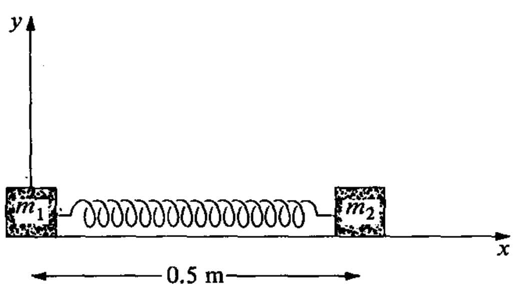


Figure 9-15

Solution Since the motion is confined to the x-axis, it suffices to consider only X_c, the x-component of the coordinates of the center of mass.

(a) In terms of the coordinate system defined in Figure 9-15, we have

$$X_c = \frac{m_1x_1 + m_2x_2}{m_1 + m_2} = \frac{0 + 2\ \text{kg} \times 0.5\ \text{m}}{1\ \text{kg} + 2\ \text{kg}}$$
$$= 0.33\ \text{m}$$

since the original position of m_1 is at the origin while m_2 is at a distance of 0.5 meter from it.

(b) Solving (9-15) for $\mathbf{V}_c$, and using the fact that all motion is along the x-axis, we find that

$$V_c = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2} = \frac{1\ \text{kg} \times 2\ \text{m/s} + 0}{1\ \text{kg} + 2\ \text{kg}}$$
$$= 0.67\ \text{m/s}$$

since m_2 is originally at rest.

(c) Since the system is isolated, so that the only forces that can influence the motion of each of the blocks is that produced by the spring, it follows that the total momentum is conserved. According to (9-13) and (9-15), this means that the velocity V_c of the center of mass is a constant, which must be equal to its initial value $V_c(0) = 0.67$ m/s. Using the fact that initially the center of mass is at the point $X_c = 0.33$ meter, it follows by use of the results from (a) and (b) that

$$X_c(t) = 0.33 + 0.67t$$

where all distances are measured in meters and t is in seconds.

This very simple motion of the center of mass of the system is to be contrasted with the motions of the individual bodies, which are exceedingly complex. For as m_1 starts to move to the right, the spring joining the two bodies becomes compressed. This leads to a force between them that tends to slow down m_1 and to accelerate m_2 to the right. Thereby the spring expands, modifying the motion even further, and so forth.

9-8 Systems of more than two particles

To generalize the ideas of momentum conservation to systems consisting of more than two particles consider, in Figure 9-16, a system of N particles. Define the momentum $\mathbf{P}$ of this collection to be the vector sum of the individual momenta:

$$\begin{aligned} \mathbf{P} &= \sum \mathbf{p}_i \\ &= \sum m_i \mathbf{v}_i \end{aligned} \tag{9-16}$$

where the sum is to be carried out over all N particles in the system, and where m_i is the mass of the ith particle and $\mathbf{v}_i$ is its velocity. We shall now establish that just as for the case of a two-particle system, if this system is isolated, so that the force experienced by each member of the system is due

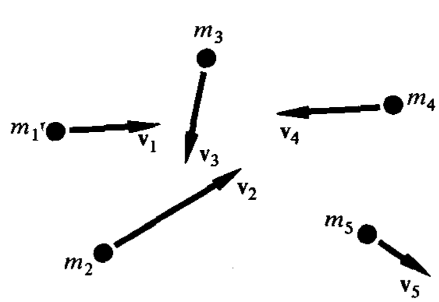


Figure 9-16

exclusively to the others, then the total momentum $\mathbf{P}$ is a constant of the motion.

To this end, let us take the derivative with respect to time of (9-16). The result is

$$\begin{aligned}\frac{d\mathbf{P}}{dt}&=\frac{d}{dt}(\mathbf{p}_1+\mathbf{p}_2+\cdots)\\&=\frac{d\mathbf{p}_1}{dt}+\frac{d\mathbf{p}_2}{dt}+\cdots\\&=\mathbf{F}_1+\mathbf{F}_2+\cdots\\&=\sum\mathbf{F}_i\end{aligned}\tag{9-17}$$

where the third equality follows by use of Newton's law of motion for each particle and where, in the last, $\mathbf{F}_i$ ($i=1,2,\ldots,N$) is the total force acting on the ith particle. Since there are no external forces acting on the system, it follows by use of Newton's third law that the sum on the right-hand side in the last equality vanishes. For the force on, say, particle number one due to particle number two is equal and opposite to that on number two due to number one. And since both these forces appear on the right-hand side of (9-17), it follows that they cancel. Similarly, it can be shown that all internal forces acting between the particles cancel in pairs. Hence it follows that

$$\frac{d\mathbf{P}}{dt}=0\tag{9-18}$$

or, in other words, the quantity $\mathbf{P}$ as defined in (9-16) is a constant of the motion. In particular, if $\mathbf{v}_1^0, \mathbf{v}_2^0, \ldots$ are the velocities of the particles of an isolated system at a fixed instant and if $\mathbf{v}_1, \mathbf{v}_2, \ldots$ are the corresponding velocities at any time t, then

$$m_1\mathbf{v}_1+m_2\mathbf{v}_2+\cdots=m_1\mathbf{v}_1^0+m_2\mathbf{v}_2^0+\cdots\tag{9-19}$$

One of the very important consequences of (9-19) is that the law of conservation of momentum is applicable also to bodies of finite dimensions. That is, although the law of momentum conservation was derived above for a system of *particles*, once the validity of this principle has been established for particles it can easily be extended to bodies of finite extension. Since a body can be thought of as consisting of a collection of point particles, it follows from the additivity property of momentum as expressed in (9-16) that the momentum of a body is the sum of the momenta of its individual members. Therefore, if in (9-19) n particles comprise one body A and the remaining $(N-n)$ comprise a second body B, then by appropriately grouping the terms in (9-19), it follows that the sum of the momenta of A and B is a constant of the motion. When viewed in this way, (9-19) is simply the law of *momentum conservation for two bodies*. Therefore, although we have derived the law of momentum conservation only for particles, because of the additivity property in (9-16) this law is valid regardless of the sizes or shapes of the bodies which comprise the system.

9-9 The principle of rocket propulsion

All methods of travel through space that we know of involve the principle of *rocket propulsion*, which in turn is based on the conservation-of-momentum law. Basically, the idea is that if you eject something from a spaceship in one direction, then the spaceship receives an acceleration in the opposite direction. This principle was illustrated in Example 9-2, in which a man standing on a smooth surface fired a pistol in a certain direction and, as a result, acquired a velocity increment in the opposite direction. In effect, the man has a momentary acceleration and, in accordance with Newton's law, he must have experienced a force.

Consider, in Figure 9-17a, a spaceship of total mass M (including that of the fuel), traveling at a velocity $\mathbf{v}$ through space. Figure 9-17b shows the situation at an instant after an amount of fuel of mass ΔM has been ejected at a velocity $\mathbf{u}$ relative to the ship. (Its velocity with respect to the given coordinate system is $(\mathbf{u} + \mathbf{v})$.) As a result of this ejection of fuel, the spaceship recoils in the opposite direction and acquires the final velocity $(\mathbf{v} + \Delta\mathbf{v})$. To apply the idea of momentum conservation, it is convenient to think of the system as consisting of two bodies, one of mass $(M - \Delta M)$ and the other of mass ΔM, both of which have the velocity $\mathbf{v}$ before the "collision" and the respective velocities $(\mathbf{v} + \Delta\mathbf{v})$ and $(\mathbf{u} + \mathbf{v})$ afterward. Substituting these values into the law of momentum conservation, we obtain

$$(M - \Delta M + \Delta M)\mathbf{v} = (M - \Delta M)(\mathbf{v} + \Delta\mathbf{v}) + \Delta M(\mathbf{u} + \mathbf{v}) \qquad \textbf{(9-20)}$$

and neglecting the term $-\Delta M\,\Delta\mathbf{v}$, which is quadratic in the small quantities and thus negligible, this simplifies to

$$\Delta\mathbf{v} = -\mathbf{u}\frac{\Delta M}{M} \qquad \textbf{(9-21)}$$

Hence the spaceship receives an increment of velocity $\Delta\mathbf{v}$, which is opposite in direction to the velocity of ejection $\mathbf{u}$ and is directly proportional to the

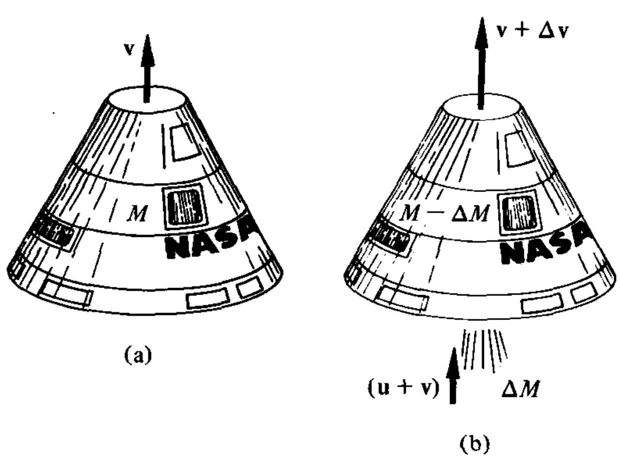


Figure 9-17

mass ΔM of the fuel ejected. By regulating the direction in which fuel is ejected from the spaceship it is possible to change the direction as well as the magnitude of its velocity, and thus to propel the ship through space.

An alternate way of viewing (9-21) may be obtained in the following way. Imagine that a space engine is operated continuously, and that an amount of fuel ΔM is ejected with a velocity $\mathbf{u}$ in a small time interval Δt. Dividing both sides of (9-21) by Δt and taking the limit $\Delta t \to 0$ we find, by use of the definition of a derivative, that this relation becomes

$$M\frac{d\mathbf{v}}{dt} = -u\frac{dM}{dt} \tag{9-22}$$

Comparing this with Newton's second law we see that, in effect, there is a force $\mathbf{F}_{th}$ or a *thrust* on the spaceship

$$\mathbf{F}_{th} = -\mathbf{u}\frac{dM}{dt} \tag{9-23}$$

and that the spaceship accelerates under the action of this force. Note that $\mathbf{F}_{th}$ is directed opposite to the ejection velocity $\mathbf{u}$ and that the larger are $|\mathbf{u}|$ and the rate dM/dt at which fuel is ejected, the larger is $\mathbf{F}_{th}$.

Another method of propulsion that operates on the same physical principle is known as *jet propulsion.* It differs from rocket propulsion in that for jet-propelled systems matter is absorbed from the environment, is then heated up by the engine, and finally ejected in a direction opposite to that in which acceleration is desired. The mass of a system driven by a jet engine does not change in time, whereas the mass of a rocket-propulsion system decreases steadily during the time that the engine runs.

†9-10 Momentum conservation of relativistic particles

The law of momentum conservation, as noted in Section 9-2, is valid even at relativistic speeds for which Newtonian mechanics is no longer applicable. The purpose of this section is to describe briefly the form of the momentum conservation law which is valid for speeds close to that of light.

Consider, in Figure 9-18a, the elastic collision of two particles of masses m_1 and m_2 with velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively. As shown in Figure 9-18b, let m_1' and m_2' and $\mathbf{v}_1'$ and $\mathbf{v}_2'$ be their respective masses and velocities after the collision. Experiment shows that a conservation-of-momentum law of the

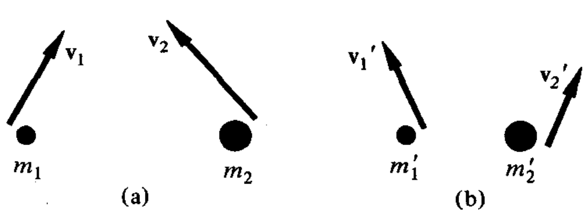


Figure 9-18

form in (9-4), that is,

$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_1' + \mathbf{p}_2' \tag{9-24}$$

is also valid, provided that the momentum $\mathbf{p}$ of a particle of mass m when it is traveling at a velocity $\mathbf{v}$ is defined by

$$\mathbf{p} = \frac{m\mathbf{v}}{[1 - v^2/c^2]^{1/2}} \tag{9-25}$$

For speeds $v \ll c$ this formula for $\mathbf{p}$ reduces to the corresponding one in (9-3). Thus (9-24), with momentum defined in terms of the velocities of the particles in (9-25), is consistent with the nonrelativistic form in (9-3) and (9-4).

For elastic collisions, experiment shows that the various momenta $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_1'$, $\mathbf{p}_2'$ must satisfy, in addition to (9-24), the relation

$$\mathscr{E}_1 + \mathscr{E}_2 = \mathscr{E}_1' + \mathscr{E}_2' \tag{9-26}$$

where, according to (7-33),

$$\mathscr{E}_1 = \frac{m_1 c^2}{[1 - v_1^2/c^2]^{1/2}} \tag{9-27}$$

with similar relations for $\mathscr{E}_2$, $\mathscr{E}_1'$, and $\mathscr{E}_2'$. In the limit of small velocities it follows from (7-34) that (9-26) reduces to the conservation-of-energy relation in (9-12), *provided* that the relation

$$m_1 + m_2 = m_1' + m_2' \tag{9-28}$$

is satisfied. Note that in our consideration of nonrelativistic collision problems we have implicitly assumed the validity of this *conservation-of-mass relation.* Indeed, in the problems it is established that (9-28), which may be thought of as a law of mass conservation, is a necessary consequence of the requirement of Galilean invariance. Relativistically, however, mass need not be conserved. Indeed, experiments show that mass and energy can be transformed into one another under appropriate circumstances and that (9-26), which does not assume the conservation of mass, is more generally valid than is the corresponding nonrelativistic form in (9-12).

Example 9-15 A particle of mass m is traveling at a velocity $\mathbf{v}_0$ when it strikes an identical particle initially at rest. If, as a result of the collision, these two particles coalesce to form a new particle, calculate its mass M.

Solution If $\mathbf{V}$ is the velocity of the final particle of mass M, then, according to the laws of conservation of momentum and of energy,

$$0 + \frac{m\mathbf{v}_0}{[1 - v_0^2/c^2]^{1/2}} = \frac{M\mathbf{V}}{[1 - V^2/c^2]^{1/2}}$$

$$\frac{mc^2}{[1 - v_0^2/c^2]^{1/2}} + mc^2 = \frac{Mc^2}{[1 - V^2/c^2]^{1/2}}$$

where use has been made (9-24) through (9-27). Eliminating the parameter **V** between these relations, we obtain

$$M^2 = 2m^2\left[1 + \frac{1}{[1 - v_0^2/c^2]^{1/2}}\right]$$

and thus the mass M of the coalesced particle exceeds the sum of the masses $2m$ of the initially colliding particles. This illustrates that mass need not be conserved under conditions involving relativistic speeds.

9-11 Summary of important formulas

The total momentum $\mathbf{P} = \Sigma \mathbf{p}_i$ of an isolated system is a constant of the motion. In particular, if $\mathbf{p}_1^0$ and $\mathbf{p}_2^0$ are the initial momenta of a two-body isolated system, then

$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_1^0 + \mathbf{p}_2^0 \qquad \textbf{(9-4)}$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the corresponding momenta at any other time. In terms of velocities, this relation is

$$m_1\mathbf{v}_1 + m_2\mathbf{v}_2 = m_1\mathbf{v}_1^0 + m_2\mathbf{v}_2^0 \qquad \textbf{(9-2)}$$

The center of mass $\mathbf{R}_c$ of a two-particle system is defined by

$$\mathbf{R}_c = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} \qquad \textbf{(9-14)}$$

and is that point along the line joining the particles that divides it in the ratio m_1/m_2. For an isolated system the center of mass travels with uniform velocity in a straight line.

The impulse **J** imparted to a body by an impulsive force is related to the change in momentum $(\mathbf{p}_f - \mathbf{p}_i)$ of that body by

$$\mathbf{J} = \mathbf{p}_f - \mathbf{p}_i \qquad \textbf{(9-7)}$$

QUESTIONS

1. Define or describe briefly what is meant by the terms (a) isolated system; (b) impulsive force; (c) impulse; and (d) inelastic collision.
2. Describe the circumstances for which the momentum of a two-body system will be conserved. By reference to a particular example, show that it is possible for one component of the total momentum of a two-body system to be conserved, even though the system is not isolated.
3. Review briefly the experimental basis for the validity of the law of momentum conservation for an isolated, two-body system.
4. Describe in detail the relation of the law of conservation of momentum to Newton's third law. If you were to observe a violation of the conservation-of-momentum law, would this imply a violation of the third law? Explain.
5. Is it possible, for an isolated two-

body system, that momentum is not conserved? If so, describe the circumstances.

6. Show, by reference to a particular example, that the momentum of an isolated two-body system may be conserved although its energy is not. Is the converse possible? That is, is it possible for the energy of an isolated two-body system to remain constant although its momentum varies?
7. Need the velocities of the two bodies of an isolated system be separately constant in time? Explain.
8. Consider, in Figure 9-19, a man walking on a heavy board, which in turn lies on a smooth sheet of ice. As the man walks to the left, why must the board start moving to the right? Under what circumstances will the man's position relative to the ice not change?

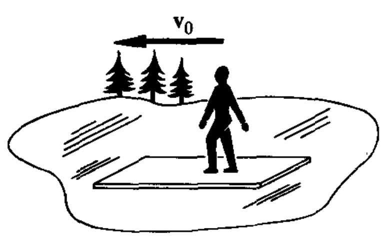


Figure 9-19

9. You observe a cart traveling at a uniform speed across a horizontal air track and watch it collide with a second cart originally at rest. Describe what measurements you must take to obtain the impulse imparted to each of the carts.
10. What is the *direction* of the impulse imparted to a pitched baseball which is hit straight back at the pitcher?
11. A pion at rest decays into a muon and a neutrino in accordance with the reaction

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

Since the pion decays from a state at rest, what can you say about the relative directions of motion of the muon and the neutrino? What could you say about their relative direction of motion if the pion were in motion when the decay took place?

12. Making use of Newton's third law, present an argument to show explicitly that the total momentum of an isolated three-particle system must be conserved.
13. Using the fact that the total momentum of an isolated *N-particle* system must be conserved, present an argument to show that the total momentum of an isolated three-*body* system must be a constant of the motion.
14. Is it necessary, in order for the law of momentum conservation to be valid, that the force between two particles lie along the line joining them? Explain.
15. Describe the sense in which the motion of the center of mass of an isolated two-particle system is governed by Newton's first law.
16. Describe, in physical terms, the physical mechanism underlying the principle of rocket propulsion.
17. Contrast and compare jet propulsion and rocket propulsion.
18. Consider two particles of mass m_1 and m_2 and separated by certain distance d. Where is the center of mass of this system if **(a)** $m_1 = m_2$? **(b)** $m_1 \ll m_2$?
19. A particle of mass m originally traveling at a velocity $\mathbf{v}_0$ strikes a second particle originally at rest. Why is it not possible for the original particle to have a velocity $-\mathbf{v}_0$ after the collision?
20. Explain in physical terms why you would expect the law of momentum conservation for an isolated two-body system to be invariant under Galilean transformations.

PROBLEMS

1. Two small bodies of respective masses $m_1 = 10$ grams and $m_2 = 25$ grams are connected by a spring and confined to one-dimensional motion. If at a certain instant the velocity v_1 of m_1 is 10 cm/s and the velocity v_2 of m_2 is 20 cm/s, find the velocity of m_2 at a subsequent instant when the velocity of m_1 has the value (a) 15 cm/s; (b) -10 cm/s.
2. Consider the two bodies in Figure 9-15 and suppose that $m_1 = 1$ kg and $m_2 = 2$ kg. Assume that at a certain instant, m_1 is at rest and m_2 has a velocity 1.5 m/s, directed to the right.
   (a) Calculate the velocity of m_1 at a subsequent instant when m_2 has the velocity 0.5 m/s, directed to the left.
   (b) How much kinetic energy has been gained or lost by the bodies?
   (c) What is the change in the potential energy of the spring?
3. Two bodies, A and B, each have a mass of 0.5 kg and are given electric charges of opposite sign so that they attract each other. If they are on a horizontal and smooth surface and, if at some fixed instant, A approaches B with a velocity of 0.1 m/s while B is stationary:
   (a) What is the velocity of B at a subsequent instant when A is traveling at a velocity 0.2 m/s toward B?
   (b) By how much has the kinetic energy of the two bodies increased or decreased in this process? What is the source of this energy?
4. Repeat both parts of Problem 3, but suppose this time the two bodies are electrically neutral and are connected by a massless spring.
5. Small magnets are attached to two carts on a horizontal air track so that they attract each other. If the mass of each cart is m, and one is initially at rest while the other approaches the first with a velocity v_0, what is the increase in the kinetic energy of each body subsequently when the speed of the first cart is $2v_0$?
6. Show that the conservation-of-momentum relation in the form in (9-2) is invariant under Galilean transformations. That is, show that if $\mathbf{v}_1^0$, $\mathbf{v}_2^0$, $\mathbf{v}_1$, and $\mathbf{v}_2$ satisfy (9-2), then so do the velocities $(\mathbf{v}_1^0 + \mathbf{u})$, $(\mathbf{v}_2^0 + \mathbf{u})$, $(\mathbf{v}_1 + \mathbf{u})$, and $(\mathbf{v}_2 + \mathbf{u})$ for any constant velocity $\mathbf{u}$. If mass were not conserved, would this requirement still be satisfied? Explain.
7. A cannon of mass 4.0×10^3 kg is mounted on wheels and fires a shell of mass 2.0 kg in a horizontal direction with a muzzle velocity of 300 m/s. With what velocity does the cannon recoil?
8. Consider again the physical system of Problem 7, but suppose this time that the cannon is aimed so that it makes an angle of 30° with respect to the horizontal. Assume the same muzzle velocity.
   (a) What is the recoil velocity of the cannon this time?
   (b) Calculate the kinetic energy acquired by the cannon and the shell, and describe the source of this energy.
9. An airplane of mass 2.0×10^4 kg is traveling horizontally at a velocity of 1.5×10^3 km/hr when the pilot fires a 2-kg shell in the forward direction. If the speed of the shell relative to the plane is 6.0×10^3 m/s:
   (a) Calculate the final velocity of the plane as seen by an ob-

server at rest with respect to the ground.

(b) Calculate the recoil velocity of the plane as seen by an observer who is originally at rest with respect to the plane.

(c) Compare your results from (a) and (b) in light of Problem 6.

10. A 50-kg boy stands on roller skates and throws, in a horizontal direction, a 200-gram baseball. If the observed recoil velocity of the boy is 3 cm/s, what was the initial speed of the baseball?

11. Consider, in Figure 9-20, a lightweight marble of mass m rolling at a speed of 20 cm/s toward a very massive metallic sphere of the same size and of mass M. Suppose that as a result of the collision the marble comes to rest.

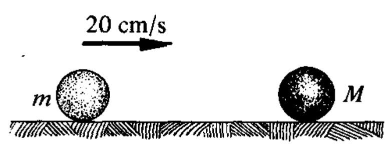


Figure 9-20

(a) What is the final velocity of the metallic sphere?

(b) Assuming that the radii of the spheres are very small, calculate the kinetic energy lost by the system in this collision.

12. Two identical carts, each of mass m, travel in opposite directions on a horizontal air track and approach each other. If one of them is traveling at a speed v_0 and the second at a speed $\frac{1}{2}v_0$ and they stick together after the collision:

(a) What is the final velocity of the composite system?

(b) How much energy has been lost? Where has this energy gone?

13. An astronaut of mass 80 kg finds himself floating in free space at a distance of 150 meters from his capsule and notes that his distance from the ship is not changing. He takes a tool from his belt and throws it in a direction away from the capsule.

(a) If it takes him 50 seconds (after throwing the tool) to return to the capsule, what is his velocity of return relative to the capsule?

(b) If the tool has a mass of 1.5 kg, with what velocity relative to the capsule was it thrown?

14. Suppose, in a lunar docking maneuver, that a landing module of mass M approaches the command ship of mass $5M$ at a relative speed of 3 km/hr. What is the resulting velocity of the system at the completion of the maneuver?

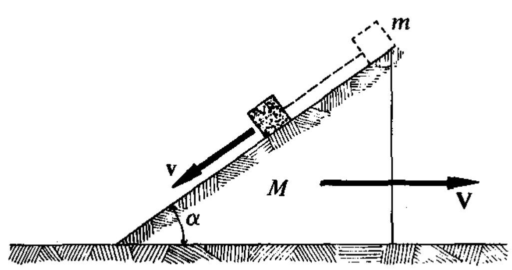


Figure 9-21

15. Consider, in Figure 9-21, an inclined plane of angle α and of mass M at rest on a horizontal and smooth surface. Suppose that a particle of mass m is placed at the top of the inclined plane and allowed to slide down. Assuming that the contact between the particle and the plane is smooth, calculate the velocity **v** of the particle relative to the inclined plane at a subsequent instant when the plane is moving to the right at a velocity **V**. (*Hint*: Why must the total momentum of this system along the horizontal direction be conserved?)

16. A small metallic sphere of mass m is dropped from a vertical distance

h above a horizontal and hard surface.

(a) If it rises to a height $h/2$ after its first bounce, calculate the impulse imparted to the sphere.

(b) If on the second bounce it rises to a height $h/4$, on the third bounce to a height $h/8$, and so forth, calculate the total impulse delivered to the surface.

17. What impulse is imparted to a plastic bag containing 2 liters of water if it strikes the ground after having fallen through a vertical distance of 20 meters?

*18. An electron, initially at rest, is subject to a uniform electric field that exerts on it a constant force of 1.6×10^{-13} newton. Suppose that the force lasts only for a time interval 10^{-9} second.

(a) Calculate the impulse imparted to the electron by evaluating the area under the $F_l(t)$ curve.

(b) From your result to (a), calculate the final velocity of the electron.

(c) Solve Newton's law for the trajectory of the electron and compare your answer to (b) with the corresponding result obtained in this alternate way.

19. A dart of mass 250 grams is thrown so that it strikes a dart board at a horizontal speed of 3 m/s. What impulse is imparted to the board if the dart sticks? What impulse would be imparted to the dart if it did not stick but dropped vertically to the ground afterward?

20. A baseball traveling at a horizontal velocity of 90 km/hr is struck by a bat. What impulse is imparted to the ball if it has a mass of 0.2 kg and: **(a)** The ball comes straight back at the pitcher with the same horizontal speed? **(b)** The ball rises vertically upward to a height of 30 meters after being struck?

21. A particle of mass m slides along the horizontal, smooth surface when it collides with a stick held rigidly in place, as shown in Figure 9-22. Assuming that the direction of its approach makes an angle θ with the normal and that it makes the same angle θ with respect to the normal after rebounding, calculate the impulse given to the rod.

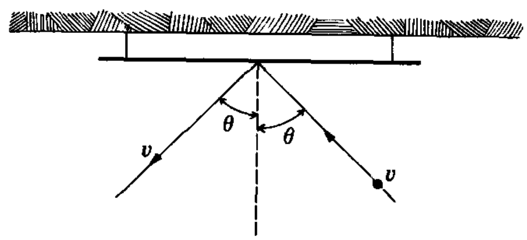


Figure 9-22

22. In an automobile collision, two autos driving at 100 km/hr in opposite directions collide head-on. If the mass of each auto is approximately 2.0×10^3 kg, what impulse is imparted to each automobile?

23. Consider the collision of two particles of mass m and M, respectively. Show by use of the result of Problem 6 that if the collision is elastic in one coordinate system, then it is also elastic in any coordinate system that moves with a uniform velocity $\mathbf{u}$ relative to the first. Is an inelastic collision inelastic in all coordinate systems?

24. Consider the elastic collision of two identical particles each of mass m. Show, that if the collision is head-on, so that the motion takes place along a single line, then the incident particle comes to rest and the target particle, assumed to be at rest initially, moves in the same direction and with the same speed that the incident particle had prior to the collision.

*25. Consider the elastic collision of two identical particles. Prove that if

before the collision one of them is at rest, then afterward either the other one is at rest or else they travel in perpendicular directions. (*Hint*: Show that if $\mathbf{v}'$ and $\mathbf{V}'$ are the velocities after the collision, then $\mathbf{v}' \cdot \mathbf{V}' = 0$.)

26. A proton traveling at a speed of 5.0×10^5 m/s collides elastically with a second proton initially at rest. One of the protons is observed coming out of the collision at an angle of 30° with respect to the incident direction.
(a) At what angle with respect to the incident direction is the other proton moving after the collision?
(b) What is the velocity of each proton after the collision?

27. Consider the elastic collision of two identical particles for the case in which one of them is initially at rest. Show that the angle of emergence with respect to the incident direction of each of the particles after the collision must be less than 90°. That is, show that the momentum of neither of the emerging particles can have a negative value for the component of its momentum along the incident direction. Is it possible for one of them to come out at right angles to this direction?

28. A 50-gram marble rolls at a speed of 0.6 m/s and strikes a second marble twice as massive as itself. If after the collision the lighter marble is observed to be traveling in a direction perpendicular to its original direction of motion at a speed of 0.2 m/s:
(a) In what direction does the more massive marble travel after the collision?
(b) What is its speed?
(c) How much energy is lost in the collision?

***29.** Consider again the collision described in Problem 26, but this time from the viewpoint of a moving observer for whom the total momentum of the two protons (before and after the collision) vanishes.
(a) Find the velocity $\mathbf{u}$ of this moving observer relative to the original system.
(b) Obtain a value for each of the four speeds in this new system.
(c) Calculate the energies of the particles as seen by this observer and show that energy is conserved in this new system as well. (*Note*: This coordinate system, in which the total momentum of the particles before and after the collision vanishes, is called the *center-of-mass system*. Sometimes it is also known as the *barycentric system*.)

30. Suppose the 1-kg ballistic pendulum in Figure 9-10 has a length of 1 meter and is struck by a 10-gram bullet, which becomes embedded in it and that the pendulum subsequently swings through an angle of 30°.
(a) What was the initial speed of the bullet?
(b) Calculate the energy *lost* in this collision.

31. A proton crashes into an α particle initially at rest. On emerging, the proton travels at an angle of 65° with respect to its incident direction, and the α particle recoils at a corresponding angle of 45° with a velocity of 2.0×10^6 m/s.
(a) Calculate the velocity of the proton after the collision.
(b) What was the proton's original energy?

32. Calculate the impulse received by the proton in Problem 31 as a result of its collision with the α particle.

*33. A small particle of mass M moving at a speed v_0 collides elastically with a lighter particle of mass m initially at rest.
   (a) Define an appropriate set of velocities and write down the implications of the laws of conservation of energy and momentum for this case.
   (b) Show that after the collision the more massive particle will be traveling with a certain velocity V in a direction making an angle θ with the incident direction and related by

$$V^2\left[1+\frac{m}{M}\right]-2Vv_0\cos\theta + v_0^2\left[1-\frac{m}{M}\right]=0$$

34. (a) Making use of the result of (b) in Problem 33, show that the largest value for the scattering angle θ of the more massive particle, call it θ_m, satisfies

$$\cos^2\theta_m = 1-\frac{m^2}{M^2}$$

   (b) Assuming that the more massive particle scatters at its maximum angle, calculate the magnitude of the velocity of the lighter particle.

35. An α particle collides elastically with a proton originally at rest. Using the result of Problem 34, calculate the maximum angle that the α particle can be deflected.

36. Calculate the center of mass of the following "two-particle" systems. In each case, the numerical value of the ratio of their masses is given by the parameter α, and their separation distance is given by R_0. Express your answer in terms of the distance from the more massive body.
   (a) Earth and moon: $\alpha = 0.012$, $R_0 = 3.84\times 10^8$ meters.
   (b) Sun and Saturn: $\alpha = 2.86\times 10^{-4}$, $R_0 = 1.42\times 10^9$ km.
   (c) Electron and proton: $\alpha = 1/1836$, $R_0 = 5.3\times 10^{-11}$ meter.

37. A man of total mass $m = 85$ kg stands on a 300-kg wooden plank, which rests on a horizontal and smooth sheet of ice. Assume that, as in Figure 9-19, the man starts to walk at a uniform speed (relative to the plank) of 1.5 m/s.
   (a) What is the velocity of the plank relative to the ice?
   (b) What is the man's velocity relative to the ground?
   (c) At what speed, relative to the platform, must he travel so that his speed relative to the ground is 1.5 m/s?

38. A 4.0×10^3-kg truck is traveling due north at 100 km/hr on a certain highway on which there is also a 2.0×10^3-kg automobile traveling due south at 70 km/hr.
   (a) At an instant when the two vehicles are 1.5 km apart and approaching each other, how far from the truck is their center of mass?
   (b) What is the velocity of their center of mass (magnitude and direction) relative to the road? What is this velocity relative to that of the automobile?

39. If the automobile in Problem 38 comes to rest but the truck continues to move in a northerly direction at a speed of 100 km/hr, what is the speed of the center of mass in this case?

40. A 10-gram particle moves along a trajectory $\mathbf{r}_1(t)$, which in a certain coordinate system is given by

$$\mathbf{r}_1(t) = 2t\,\mathbf{i} + t^3\mathbf{j}$$

where t is measured in seconds and the distance is measured in cen-

timeters. Suppose that there is a second particle, of mass 15 grams, moving along the trajectory

$$\mathbf{r}_2(t) = (3 + 4t^2)\mathbf{i} - t^3\mathbf{j}$$

where all quantities have been expressed in the same coordinate system and units as above.

(a) Write down $\mathbf{R}_c(t)$, the trajectory of the center of mass of the two particles at any time t.
(b) Calculate the velocity $\mathbf{V}_c$ of the center of mass of the system.
(c) Is the momentum of this two-particle system conserved along any direction?

41. Consider a system of three particles of mass m_1, m_2, and m_3. If $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ are the coordinates of these particles at any time t, show that if they are isolated, then the time derivative of the quantity

$$m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + m_3\mathbf{r}_3$$

is constant in time. Can you generalize this result to a system of N particles?

42. Consider the two bodies in Figure 9-15 with $m_1 = 1$ kg and $m_2 = 2$ kg. Suppose that the bodies are on a smooth and horizontal surface and start out initially from a state of rest. If m_1 suddenly receives an impulsive blow corresponding to an impulse $J = 1.5$ N-s directed to the right in the figure:

(a) What is the velocity of m_1 immediately afterward? What is the velocity of m_2?
(b) What is the velocity of the center of mass immediately after the impulse has been delivered?
(c) Write down the position of the center of mass of this two-body system at any time t afterward.

43. Consider again the physical situation described in Problem 42. Suppose that this time m_2 is struck impulsively in such a way that the center of mass initially travels to the left at a velocity of 1.5 m/s.

(a) What was the initial velocity of m_2?
(b) What was the magnitude and direction of the impulse delivered?
(c) What is the velocity of m_1 at a subsequent instant when m_2 is at rest?

†44. Show that the relation between the energy $\mathscr{E}$ and the momentum $\mathbf{p}$ of a relativistic particle of mass m is

$$\mathscr{E} = [c^2\mathbf{p}^2 + m^2c^4]^{1/2}$$

†45. Consider the decay at rest of a particle of mass M into two particles of mass m_1 and m_2. Show that the energies $\mathscr{E}_1$ and $\mathscr{E}_2$ of these two particles, respectively, have the values

$$\mathscr{E}_1 = \frac{M^2 + m_1^2 - m_2^2}{2M}c^2$$

and a similar formula for $\mathscr{E}_2$.

†46. Making use of the results from Problem 45 and the table of masses in Chapter 1, calculate the energies of:

(a) The muon and the neutrino in the reaction

$$\pi^+ \to \mu^+ + \nu_\mu$$

(b) The π^+ and the π° in the reaction

$$K^+ \to \pi^+ + \pi^\circ$$

10 Angular momentum and planetary motion

There are seven windows in the head, two nostrils, two eyes, two ears and... From which and many other similar phenomena of nature such as the seven metals etc. which it were tedious to enumerate, we gather that the number of planets is necessarily seven.

FRANCISCO SIZZI (Commenting on Galileo's discovery of Jupiter's moons)

I write my book, whether it be read by the present age or posterity imports little, ... has not G-d waited six thousand years for an observer of his works.

JOHANNES KEPLER

10-1 Introduction

The main purpose of this chapter is to introduce a new dynamical quantity known as *angular momentum.* Under appropriate conditions we find that, just as for the analogous cases of energy and of linear momentum, the angular momentum of certain physical systems is a constant of the motion. Associated with these systems, then, there is a law of conservation of angular momentum. However, unlike energy, which is a scalar, angular momentum is similar to linear momentum in that it is a vector. This implies that for those physical systems for which the angular momentum is

conserved, there are in general three conservation laws: one for each component of this vector.

From an overall point of view, some of the most important applications of the law of angular momentum conservation are to composite or multiparticle systems, and several of these applications will be discussed in detail in Chapters 11 and 12. Partially in preparation for these studies, in this chapter we shall define angular momentum and describe its important properties in the simpler context of a single particle. The chief application of these ideas will be directed mainly to the case of a body orbiting in the inverse-square gravitational field of a star or planet. Of particular interest in this connection are the motions of planets, comets, and other celestial bodies—including man-made ones—in our solar system. For these, as we shall see, the angular momentum about the sun is conserved. Since the energy of such a body is also conserved, it will follow that planetary motion is characterized by the two conservation laws of energy and of angular momentum. Moreover, these two constants of the motion not only characterize planetary motion but, as will be established in this chapter, they also determine all of the essential features of this motion.

We shall begin this chapter with a brief statement of Kepler's three laws of planetary motion and some of the empirical data that these laws summarize. The following three sections are then devoted to a discussion of the angular momentum of a particle and a related quantity known as *torque.* The remainder of the chapter then deals with the derivation of Kepler's laws from Newtonian mechanics.

10-2 Kepler's laws

Since the dawn of recorded history, and perhaps even before, man has observed and followed with considerable interest the motions of the stars, the planets and other visible celestial objects. By the early part of the seventeenth century, when Johannes Kepler (1571–1630) appeared on the scene, there existed thus a large body of precise astronomical data. Noteworthy among these measurements were those made by Tycho Brahe (1546–1601), the precision of whose observations was at that time second to none. Kepler acquired Brahe's data shortly after the latter's death in 1601 and then spent almost a decade analyzing them. In the year 1609 he published his *Astronomia Nova* (*New Astronomy*) in which he enunciated for the first time those rules of planetary motion that today we call Kepler's first two laws. Twenty years later, in 1629, he published his *De Harmonice Mundi* (*World Harmony*), which contains his third law.

Kepler's laws, as we know today, characterize correctly the motion not only of the planets about the sun and the moons about their host planets but the motions of man-made satellites about gravitating bodies as well. Thus the motion of the moon about the earth as well as the motion of the Apollo 11

command module about the moon, during which the photograph in Figure 10-1 was taken, are fully in accord with Kepler's laws. Specialized to the case of the sun's planets, these laws state:

1. Each planet moves about the sun in an elliptical path, with the sun at a focal point.
2. A line drawn from the sun to a planet sweeps out equal areas in equal times.
3. The ratio of the square of the period of a planet to the cube of its semimajor axis is a universal constant (in the solar system) and is thus the same for each planet.

Note that implied by the first law is the fact that the orbit of each planet must lie in a plane.

To better understand these laws, let us first review briefly some of the geometric properties of an ellipse. An ellipse may be defined as the locus of points, the sum of whose distances from two fixed points—called the foci of the ellipse—is constant. One important parameter of the ellipse is the length of its *semimajor* axis a, which is defined as the longest line to the ellipse from its center. A second is the *eccentricity* ϵ, which is defined so that $2a\epsilon$ is the

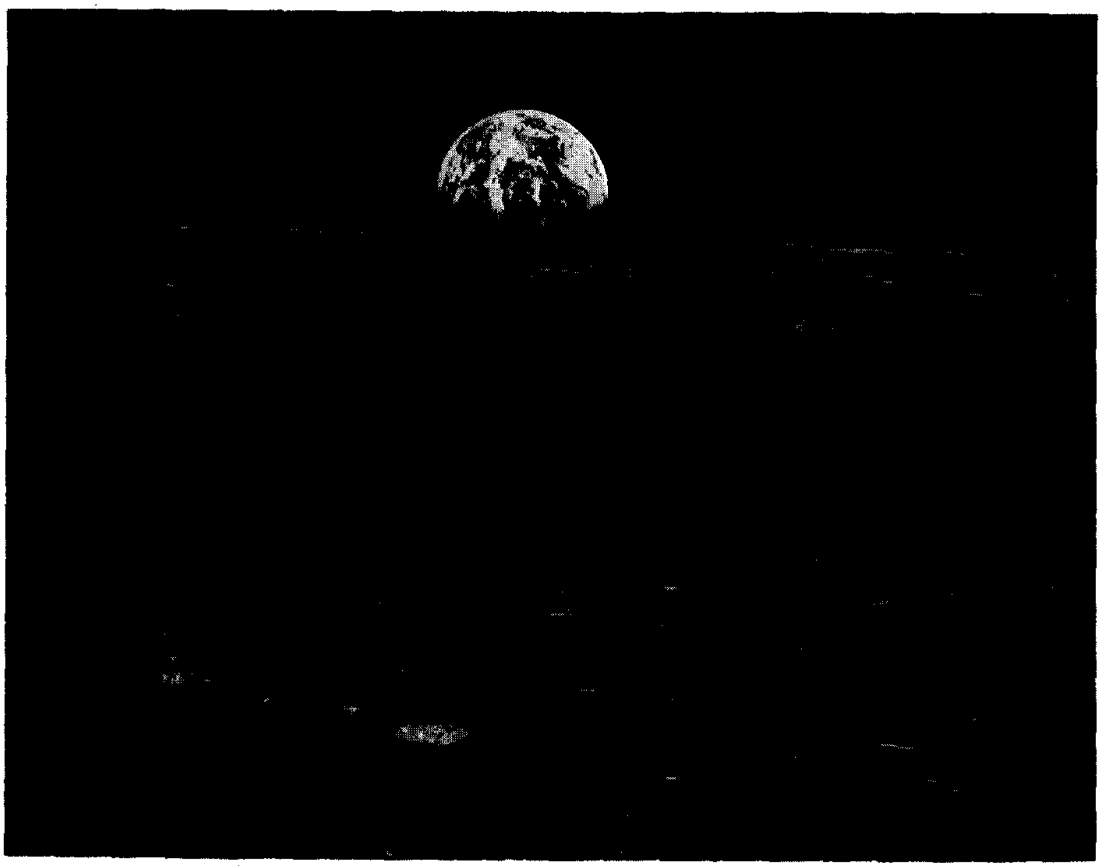


Figure 10-1 *The rising earth as seen by the Apollo 11 astronauts as they came from behind the moon just after lunar orbit insertion burn. (Courtesy of NASA.)*

distance between the foci. Figure 10-2 shows an ellipse centered at the origin of a coordinate system and with its semimajor axis parallel to the x-axis. The equation for the ellipse in this system is

$$\frac{x^2}{a^2}+\frac{y^2}{a^2(1-\epsilon^2)}=1 \tag{10-1}$$

Note that for the special case $\epsilon = 0$ the foci coincide, and consistent with Figure 10-2, (10-1) reduces to the equation of a circle of radius a. The quantity $b = a\sqrt{1-\epsilon^2}$ is also known as the *semiminor* axis of the ellipse, and its geometrical significance is illustrated in the figure.

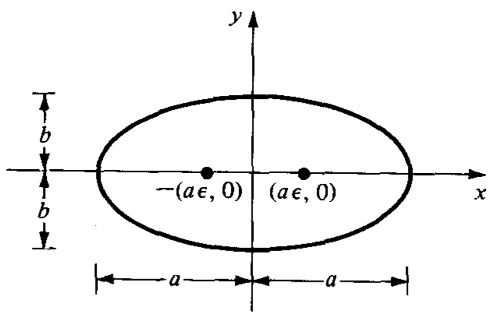


Figure 10-2

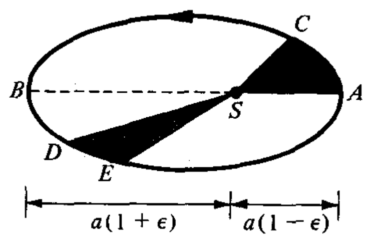


Figure 10-3

Consider now in Figure 10-3 a planetary ellipse and let S be that focal point where the sun is located. According to the first law this ellipse is a possible path for a planet, so to complete the description of its orbit it is necessary to specify the values for the length of its semimajor axis a and its eccentricity ϵ. Table 10-1 lists values for these two parameters for the nine planets in our solar system. (See also Appendix B in this connection.) The second column presents the observed values for the semimajor axis of each planet in terms of the *astronomical unit* (AU), which is defined to be the mean distance of the earth from the sun. It has the value 1.496×10^8 km. The

Table 10-1 Planetary data

Planet	*Semimajor Axis a* (AU)	*Eccentricity* ϵ	*Period T* (seconds)	a^3/T^2([AU]^3/s^2)
Mercury	0.39	0.21	7.60×10^6	1.02×10^{-15}
Venus	0.72	0.0068	1.93×10^7	1.00×10^{-15}
Earth	1.00	0.017	3.16×10^7	1.00×10^{-15}
Mars	1.52	0.093	5.92×10^7	1.00×10^{-15}
Jupiter	5.20	0.048	3.74×10^8	1.03×10^{-15}
Saturn	9.54	0.056	9.30×10^8	1.00×10^{-15}
Uranus	19.2	0.047	2.66×10^9	1.00×10^{-15}
Neptune	30.1	0.0086	5.20×10^9	1.00×10^{-15}
Pluto	39.5	0.25	7.85×10^9	1.00×10^{-15}

length of the semimajor axis of Mars, for example, is

$$1.52 \text{ AU} = 1.52 \times 1.496 \times 10^8 \text{ km} = 2.26 \times 10^8 \text{ km}$$

The third column in the table lists the values for the eccentricities of the planetary ellipses. Except for Mercury, with $\epsilon = 0.21$, and Pluto, with its large eccentricity of 0.25, the eccentricities of all the planets are very small. This implies, according to Figure 10-2, that the orbits in these cases are approximately circles. For example, since the eccentricity of the earth is 0.017, it follows by use of (10-1) that the ratio of its semiminor axis $a\sqrt{1-\epsilon^2}$ and its semimajor axis a is

$$\begin{aligned}\sqrt{(1-\epsilon^2)} &\cong \left(1 - \frac{1}{2}\epsilon^2\right) \\ &= 0.99986\end{aligned}$$

Thus, to a high degree of precision, the earth's orbit is a circle.

Let us now consider Kepler's second law. Suppose that in Figure 10-3 in a certain time interval Δt the earth goes from a point A on its orbit to a second point C. At an arbitrary subsequent instant suppose that during precisely the same time interval Δt the earth goes from D to point E. According to the second law, the two shaded areas in the figure, namely SAC and SDE, will have precisely the same values. In other words, the speed of a planet about the sun is constantly changing, but its variation is such that this rule of equal areas holds.

Column 4 of Table 10-1 lists the observed values in seconds for the periods of the planets. (Recall that the period T of a planet is the time it takes for it to make a complete orbit about the sun. In terms of Figure 10-3, for example, the period of the earth about the sun is the time it takes the earth to go from the point A completely around the elliptical path through the points $CBDE$ back to the point A.) Now, according to Kepler's third law, the ratio of the square of the period T of a planet and the cube of its semimajor axis a must be the same for all planets in the solar system. Column 5 of the table lists the measured values for the ratio a^3/T^2. The agreement between Kepler's predictions and the observed facts is truly remarkable.

It deserves to be emphasized that Kepler arrived at his laws in a purely empirical way by carefully scrutinizing and correlating the astronomical data available to him. The discovery of the fact that the motions of the planets are governed by such precise and simple rules is truly a most remarkable feat and was destined to play a decisive role in Newton's even more remarkable extrapolation of the laws of motion and gravitation.

Example 10-1 Assuming that the moon orbits the earth in a circle of radius $a_M = 3.84 \times 10^5$ km with a period $T_M = 27.3$ days, what would be the semimajor axis of a satellite which orbits about the earth with a period $T_s = 3$ hr?

Solution According to Kepler's third law, as applied to these two satellites of the earth, the ratio of the cube of the semimajor axis and the square of the period must be the same. Thus

$$\frac{a_s^3}{T_s^2} = \frac{a_M^3}{T_M^2}$$

where a_s is the length of the semimajor axis of the satellite. Solving for a_s and inserting the known values for the remaining factors, we obtain

$$a_s = a_M \left(\frac{T_s}{T_M}\right)^{2/3} = 3.84 \times 10^5 \text{ km} \left(\frac{3 \text{ hr}}{27.3 \text{ days}} \times \frac{1 \text{ day}}{24 \text{ hr}}\right)^{2/3}$$
$$= 1.1 \times 10^4 \text{ km}$$

This corresponds to a distance nearly twice the radius $R_0 = 6.4 \times 10^3$ km of the earth.

10-3 The cross product

As a preliminary to our defining angular momentum, in this section we shall describe a way of multiplying together two vectors to produce a third vector. It is known as the *cross product* of two vectors and is to be contrasted with the *dot product* which, as defined in Section 7-4, is a scalar.

Given two vectors **A** and **B**, we associate with them a third *vector*, which is known as their *cross product* or their *vector product*, and for which the symbol $\mathbf{A} \times \mathbf{B}$ (read "*A* cross *B*") is reserved. The magnitude $|\mathbf{A} \times \mathbf{B}|$ of the cross product is defined by

$$|\mathbf{A} \times \mathbf{B}| = AB \sin \theta \tag{10-2}$$

where θ is the angle between **A** and **B**. As shown in Figure 10-4 the quantity $A \sin \theta$ is that component of **A** which is perpendicular to **B**, and $B \sin \theta$ is the component of **B** perpendicular to **A**. Thus the magnitude of the vector $\mathbf{A} \times \mathbf{B}$ can be viewed as the product of the magnitude of either vector and the component of the second *perpendicular to the other.* It follows that the cross product of two parallel vectors vanishes. Also, since the area of a triangle is one half the product of its altitude and base, it follows by reference to Figure 10-5 that the area of the triangle, two of whose sides are the vectors **A** and **B**, is the product $\frac{1}{2}AB \sin \theta$. Comparison with (10-2) shows then that the cross product of the two vectors **A** and **B** is a vector whose magnitude is twice the

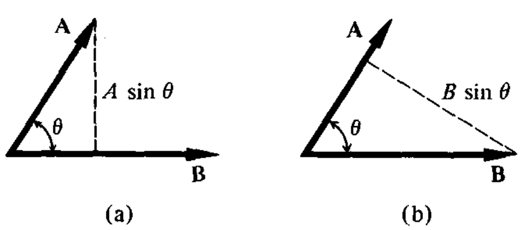


Figure 10-4

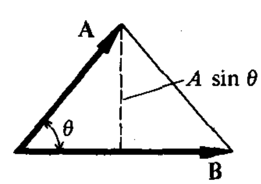


Figure 10-5

area of a triangle formed with these two vectors as sides. Equivalently, the magnitude of the cross product of two vectors can be thought as the area of a parallelogram with these two vectors as adjacent sides.

To complete the specification of the vector $\mathbf{A} \times \mathbf{B}$ it is necessary to define the direction of this vector. Since the cross product of parallel vectors vanishes, it suffices to assume that $\mathbf{A}$ and $\mathbf{B}$ are not parallel and hence determine a plane. The direction of $\mathbf{A} \times \mathbf{B}$ is defined to be perpendicular to this plane and with sense given by the *right-hand rule.* Imagine a right-handed screw with its axis oriented perpendicular to the plane determined by $\mathbf{A}$ and $\mathbf{B}$ (see Figure 10-6). If this screw is rotated so as to turn it from the direction of $\mathbf{A}$ to $\mathbf{B}$ (through the angle $\theta < \pi$), then the direction of advance of the screw is the same as the sense of $\mathbf{A} \times \mathbf{B}$. An alternate way of defining this sense of direction is shown in Figure 10-7. Imagine seizing with your right hand an axis perpendicular to the plane determined by $\mathbf{A}$ and $\mathbf{B}$ with your fingers pointing in the direction in which $\mathbf{A}$ must be pushed (through the smaller angle) to bring it into coincidence with $\mathbf{B}$. Your thumb will then point in the direction of the vector $\mathbf{A} \times \mathbf{B}$.

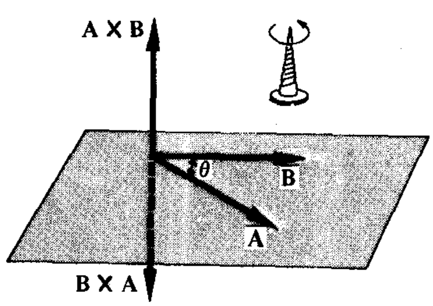


Figure 10-6

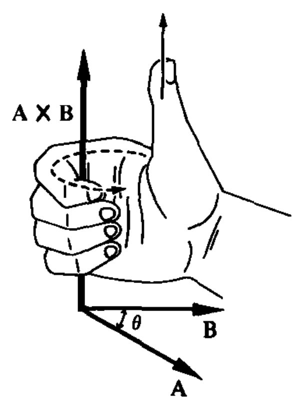


Figure 10-7

Note that there is an asymmetry in the definition of a cross product with respect to the roles played by the first and second vectors. Thus, if in Figure 10-6 the direction of the vector $\mathbf{A} \times \mathbf{B}$ is up, then the direction of the vector $\mathbf{B} \times \mathbf{A}$ is down. On the other hand, since according to (10-2) the magnitudes of the two vectors $\mathbf{A} \times \mathbf{B}$ and $\mathbf{B} \times \mathbf{A}$ are the same, it follows that the cross product satisfies the relation

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \tag{10-3}$$

This asymmetry inherent in the definition of the cross product is to be contrasted with the commutative property of the scalar product, as reflected in (7-5).

Provided that the minus sign in (10-3) is included each time the order of two factors in a cross product is exchanged, the algebra associated with the cross product is essentially the same as that involved in the dot product. In

particular, the cross product satisfies the *distributive law*

$$\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C} \tag{10-4}$$

whose similarity with (7-6) should be noted.

By way of illustration consider, in Figure 10-8, three mutually orthogonal unit vectors **i**, **j**, and **k** parallel to the respective x-, y-, and z-axes of a Cartesian coordinate system. First, since the cross product of parallel vectors vanishes, it follows that

$$\mathbf{i} \times \mathbf{i} = 0 \qquad \mathbf{j} \times \mathbf{j} = 0 \qquad \mathbf{k} \times \mathbf{k} = 0$$

To determine the cross product of the other combinations of vectors consider the vector $\mathbf{i} \times \mathbf{j}$. Since **i** and **j** are unit vectors and at right angles to each other, it follows from (10-2) that $\mathbf{i} \times \mathbf{j}$ must also be a unit vector.

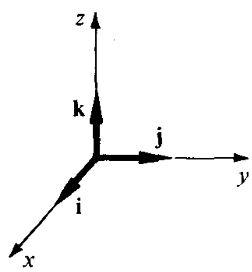


Figure 10-8

Moreover, applying the right-hand rule we see, by reference to Figure 10-8, that $\mathbf{i} \times \mathbf{j}$ must be parallel to the unit vector **k** along the z-axis. Hence

$$\mathbf{i} \times \mathbf{j} = \mathbf{k} \tag{10-5}$$

and similarly we obtain

$$\mathbf{j} \times \mathbf{k} = \mathbf{i} \qquad \mathbf{k} \times \mathbf{i} = \mathbf{j}$$

Note the order of the factors in these relations. The cross product $\mathbf{j} \times \mathbf{i}$ is antiparallel to **k** and thus is in the same direction as is the vector $-\mathbf{k}$.

Example 10-2 Calculate the cross product $\mathbf{A} \times \mathbf{B}$, assuming that **A** and **B** both lie in the x-y plane and have the respective nonzero components A_x, A_y and B_x, B_y.

Solution Making use of the distributive law in (10-4), we have

$$\begin{aligned}\mathbf{A} \times \mathbf{B} &= (\mathbf{i}A_x + \mathbf{j}A_y) \times (\mathbf{i}B_x + \mathbf{j}B_y) \\ &= \mathbf{i} \times \mathbf{j} A_x B_y + \mathbf{j} \times \mathbf{i} A_y B_x\end{aligned}$$

where the second equality follows from the fact that the cross product of parallel vectors vanishes. Making use of (10-3) and (10-5) we find that

$$\mathbf{A} \times \mathbf{B} = \mathbf{k}(A_x B_y - A_y B_x)$$

so that $\mathbf{A} \times \mathbf{B}$ points along the z-axis and is thus perpendicular to the plane containing **A** and **B**. It is left as an exercise to confirm that this formula is consistent with (10-2).

10-4 Angular momentum

Consider in Figure 10-9 a particle of mass m moving along a trajectory AB and let $\mathbf{r}$ be its position at time t relative to a certain origin O. We define the angular momentum $\mathbf{L}$ of the particle with respect to the origin O by

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \tag{10-6}$$

where $\mathbf{p}$ is its linear momentum. According to the definition of $\mathbf{p}$ in (9-3), this may be expressed equivalently as

$$\mathbf{L} = m\mathbf{r} \times \mathbf{v} \tag{10-7}$$

with $\mathbf{v}$ the instantaneous velocity of the particle. The angular momentum $\mathbf{L}$ of a particle is also referred to as its *moment of momentum.*

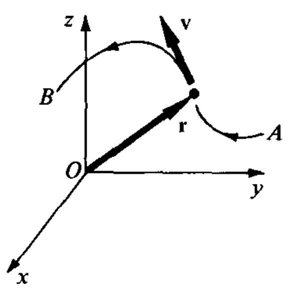


Figure 10-9

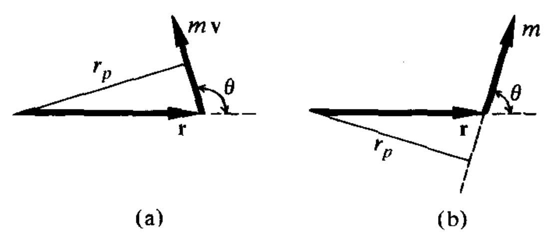


Figure 10-10

According to this definition, the magnitude of the angular momentum $\mathbf{L}$ is the product of the magnitude of the linear momentum, $m\mathbf{v}$, and the perpendicular distance, r_p, from the origin to the velocity vector or its extension (see Figure 10-10). The perpendicular distance r_p is also called the *moment arm.* The direction of the angular momentum is perpendicular to the plane defined by $\mathbf{r}$ and $\mathbf{p}$ and with a sense given in accordance with the right-hand rule. If, for example, a particle is confined to motion in a plane with its position and velocity vectors related as in Figure 10-10a, then the direction of the angular momentum $\mathbf{L}$ is perpendicular to and up out of this plane.

It follows by use of (10-7) that the unit of angular momentum is the same as the product of the units of energy and time. Hence the unit of angular momentum is the joule second (J-s).

It should be carefully noted that the angular momentum of a particle depends, in general, on the origin with respect to which it is defined. The angular momentum of a particle with respect to different origins will, in general, be different; or, equivalently, the angular momenta with respect to the same origin of two identical particles moving with precisely the same velocity but along distinct trajectories are generally different. Therefore, in

discussions involving the angular momentum **L** of a particle, the origin with respect to which this angular momentum is defined must be clearly specified.

Example 10-3 Calculate the angular momentum of the earth with respect to an origin taken at the center of the sun. Assume that the orbit of the earth is a circle of radius 1.5×10^8 km, that its speed is a uniform 30 km/s and that the earth's rotational motion about its own axis can be neglected.

Solution Since the orbit of the earth is assumed to be a circle, it follows that its velocity $\mathbf{v}$ is always perpendicular to the displacement $\mathbf{r}$ from the sun. According to (10-7) and the definition of a cross product, it follows that the angular momentum of the earth, $\mathbf{L}$, will be perpendicular to and come out of the plane of the diagram in Figure 10-11. The magnitude of this angular momentum is

$$\begin{aligned} L &= Mvr \\ &= (6.0 \times 10^{24}\ \text{kg}) \times (3.0 \times 10^4\ \text{m/s}) \times (1.5 \times 10^{11}\ \text{m}) \\ &= 2.7 \times 10^{40}\ \text{J-s} \end{aligned}$$

where the second equality follows since $M = 6.0 \times 10^{24}$ kg.

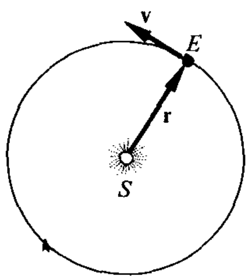


Figure 10-11

Example 10-4 Show that if $\mathbf{r}$, the instantaneous position of a particle of mass m with respect to a certain origin O, is parallel to its acceleration $d\mathbf{v}/dt$, then the time rate of change of the angular momentum $\mathbf{L}$ with respect to the given origin vanishes.

Solution Making use of the definition of $\mathbf{L}$ in (10-7), we find that

$$\begin{aligned} \frac{d\mathbf{L}}{dt} &= \frac{d}{dt}(m\mathbf{r} \times \mathbf{v}) \\ &= \frac{d\mathbf{r}}{dt} \times m\mathbf{v} + \mathbf{r} \times m\frac{d\mathbf{v}}{dt} \end{aligned}$$

where the second equality follows from the rule for differentiating a product, keeping in mind that the order of the factors must be maintained. Since $\mathbf{v} = d\mathbf{r}/dt$ and since the cross product of the parallel vectors $\mathbf{v}$ and $m\mathbf{v}$ vanishes, it follows that the first term after the second equality in the above relation is zero. Moreover, since by hypothesis $\mathbf{r}$ is parallel to $d\mathbf{v}/dt$, so is the second term. Hence

$$\frac{d\mathbf{L}}{dt} = 0$$

10-5 Torque

Consider, in Figure 10-12, a small body of mass m moving under the action of a certain force **F**, and let **r** represent its instantaneous position with respect to an origin O. We define the *torque* $\boldsymbol{\tau}$ that the force **F** exerts on the body about O by

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \tag{10-8}$$

so that the direction of this torque is perpendicular to the plane determined by **r** and **F**. Its magnitude is the product of $|\mathbf{F}|$ and the perpendicular distance r_p from the origin to the line of action of the force. This distance r_p is also known as the *moment arm* of the force, and the magnitude τ of the torque is often characterized as the *moment of the force* about the given origin. In common with angular momentum, torque also depends on the choice of origin; in general, a given force will produce a variety of torques about different origins.

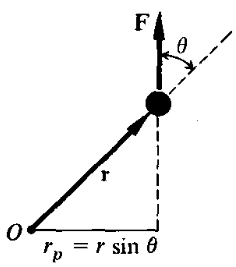


Figure 10-12

The unit of torque is the product of the unit of force and of distance; that is, the newton meter (N-m). This is the same as the unit of energy, the joule, and therefore the joule is also the unit of torque.

There is a very important relation between the torque $\boldsymbol{\tau}$ exerted on a particle by a force **F** and its angular momentum. This relation is:

The time rate of change of the angular momentum of a particle measured with respect to a Newtonian origin O is equal to the torque—defined with respect to the same origin—acting on it:

$$\boldsymbol{\tau} = \frac{d\mathbf{L}}{dt} \tag{10-9}$$

Note carefully that in (10-9) $\boldsymbol{\tau}$ and **L** are both defined with respect to the same origin O, which moreover must be Newtonian.[1] If either of these conditions is not satisfied, then (10-9) is simply *not* true.

[1]Equation (10-9) is also trivially valid if $\boldsymbol{\tau}$ and **L** are defined with respect to the accelerating origin fixed to the particle itself. For in that case **r** vanishes, and therefore so will **L** and $\boldsymbol{\tau}$, according to (10-7) and (10-8).

It is interesting to note that even if there is a force acting on a particle, the associated torque may vanish nevertheless—for example, because the line of action of the force goes through the origin so that the moment arm r_p in Figure 10-12 vanishes. In this case, (10-9) implies that the angular momentum **L** defined with respect to the given origin is a constant of the motion. This is precisely the situation, as we shall see, for planetary motion.

To establish the validity of (10-9), consider a particle of mass m under the action of a force **F**. Let **r** be the particle's position relative to a Newtonian origin O and **v** its velocity relative to the same origin. Taking the time derivative of the angular momentum **L** relative to O we find, by use of the same arguments as in Example 10-4, that

$$\frac{d\mathbf{L}}{dt} = \frac{d}{dt}(m\mathbf{r}\times\mathbf{v}) = \mathbf{r}\times m\frac{d\mathbf{v}}{dt}$$

where the term $\mathbf{v}\times m\mathbf{v}$ vanishes, since it is the cross product of parallel vectors. Since the origin is Newtonian, the factor $m\,d\mathbf{v}/dt$ is equal to the force **F** acting on the particle. With this substitution, the validity of (10-9) then follows from the definition of torque in (10-8).

Note that if the origin were not Newtonian, then the factor $m\,d\mathbf{v}/dt$ could not be replaced by the force **F** on the particle and (10-9) would *not* result in that case.

Example 10-5 A bead of mass m slides around the rim of a rough, circular wire of radius b with speed v, given by

$$v = v_0 - \alpha t$$

where v_0 and α are two positive constants (Figure 10-13). Calculate:

(a) The angular momentum of the bead relative to the center of the ring and its time derivative at time t.

(b) The torque about the center acting on the bead.

(c) The frictional force on the bead.

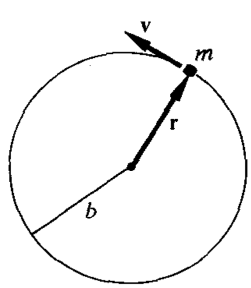


Figure 10-13

Solution

(a) Reference to Figure 10-13 shows that **r** and **v** are always perpendicular to each other and that the direction $\mathbf{r}\times\mathbf{v}$ is perpendicular to and up out of the plane of the diagram. Hence, since $|\mathbf{r}| = b$ and $|\mathbf{v}| = (v_0 - \alpha t)$,

$$\mathbf{L} = mb(v_0 - \alpha t)\mathbf{k}$$

where $\mathbf{k}$ is a unit vector perpendicular to and out of the plane of the diagram in Figure 10-13. The time derivative of $\mathbf{L}$ is

$$\frac{d\mathbf{L}}{dt} = -mb\alpha\mathbf{k}$$

(b) Substitution of (a) into (10-9) leads to

$$\boldsymbol{\tau} = -mb\alpha\mathbf{k}$$

(c) The force $\mathbf{F}$ which produces this torque must be directed opposite to the velocity vector $\mathbf{v}$ in Figure 10-13. Its magnitude is

$$F = m\alpha$$

since the moment arm of this torque is the radius b.

10-6 Angular momentum and planetary motion

Let us now turn to a study of the motion of the planets in the gravitational field of the sun with the ultimate goal of deriving Kepler's laws. Although the discussion will be centered mainly around the motion of a planet, in particular our own earth, it is to be understood that corresponding results are applicable for the motion of any satellite about any large gravitational body.

To describe the motion of a planet about the sun, let us define $\mathbf{r}$ to be the instantaneous position vector of the planet with respect to an origin at the center of the sun (see Figure 10-14). For the cases of physical interest, the separation distances of the planets from the sun are so enormous that the radii of the planets and of the sun are negligible by comparison. According to the law of universal gravitation, the force $\mathbf{F}$ on a planet is

$$\mathbf{F} = -\frac{GMm}{r^3}\mathbf{r} \tag{10-10}$$

where M is the mass of the sun and m that of the planet of interest. The minus sign follows since $\mathbf{r}$ is directed radially outward. Note the important feature that the direction of the force $\mathbf{F}$ in (10-10) is parallel (more accurately antiparallel) to the planet's position $\mathbf{r}$ and that the magnitude of the force

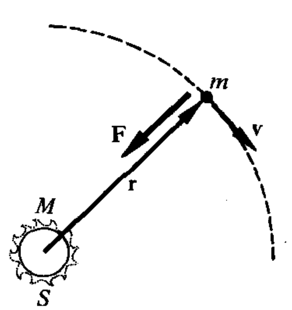


Figure 10-14

depends only on the distance r from the planet. A force of this type, which is parallel to $\mathbf{r}$ and varies only with its magnitude, is known as a *central force.* Although for the present, only the force $\mathbf{F}$ in (10-10) will be used, it should be kept in mind that many of the conclusions reached in the next section follow equally well for any central force field.

One of the very important consequences of the fact that the force $\mathbf{F}$ in (10-10) is central and is thus parallel to the direction of $\mathbf{r}$ is that the *angular momentum of the planet about the sun is conserved.* This follows from the fact that the cross product of parallel vectors vanishes, and thus the torque $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$ about the given origin is zero. Substituting the value zero for the torque $\boldsymbol{\tau}$ into (10-9), we find that the time derivative of the angular momentum vanishes. Hence the angular momentum of the planet is a constant of the motion. If, for example, a planet moves along the dashed trajectory in Figure 10-15 and if at some initial instant its position and velocity are $\mathbf{r}_0$ and $\mathbf{v}_0$, respectively, then the corresponding quantities $\mathbf{r}$ and $\mathbf{v}$ at an arbitrary subsequent instant satisfy

$$m\mathbf{r}_0 \times \mathbf{v}_0 = m\mathbf{r} \times \mathbf{v}$$

Because of this conservation of angular momentum for planetary motion, the defining relation for angular momentum in (10-7) takes on a new meaning. On the one hand, $\mathbf{L}$ is the vector function of position and velocity defined by this relation; on the other hand, however, $\mathbf{L}$ is a constant vector—that is, a vector whose magnitude and direction are fixed in time. Thus as a planet moves about the sun, even though its position $\mathbf{r}$ and velocity $\mathbf{v}$ are constantly changing their cross product in (10-7) does not.

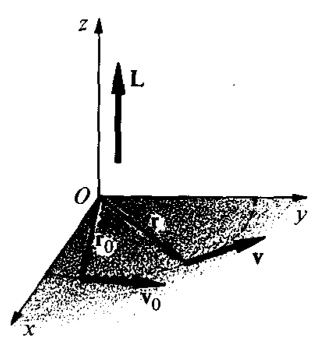


Figure 10-15

One immediate and very important consequence of the conservation of angular momentum for planetary motion is that—consistent with Kepler's first law and the observed planetary orbits—the motion of each planet is permanently confined to a single plane. To see this, suppose that at some instant, $\mathbf{r}_0$ and $\mathbf{v}_0$ are the position and velocity of the planet. Define a Cartesian coordinate system so that these two vectors lie in the x-y plane. It follows, as shown in Figure 10-15, that instantaneously the angular momentum $\mathbf{L}$ will be along the z-axis in this system. Moreover, since $\mathbf{L}$ is

conserved the angular momentum of the planet will be directed along the z-axis for all times t. In particular, if at a subsequent instant, $\mathbf{r}$ and $\mathbf{v}$ are the position and velocity of the planet, then since the angular momentum of the planet must still be directed along the z-axis, it follows that $\mathbf{r}$ and $\mathbf{v}$ must still be in the x-y plane. For if one of these two vectors had a component along the z-axis, then, by definition of a cross product, $\mathbf{L}$ would of necessity have a component lying in the x-y plane, and this would contradict the conservation of angular momentum.

This fact that the orbit of a planet must lie in a plane can also be derived more directly by use of physical arguments. Suppose in Figure 10-15 that $\mathbf{r}$ is the instantaneous position of a planet at the instant when its velocity is $\mathbf{v}$. If there were no force acting on the planet, it would continue to travel in the plane defined by $\mathbf{r}$ and $\mathbf{v}$. Moreover, if the planet is to move out of this plane, a force with a component perpendicular to the plane is required. But the gravitational force $\mathbf{F}$ in (10-10) is parallel to $\mathbf{r}$ and therefore lies entirely in this plane. Thus it follows, without direct reference to angular momentum, that the motion of the planet is permanently confined to the plane defined by the vectors $\mathbf{r}$ and $\mathbf{v}$ at any one instant. In the following we shall invariably take this to be the x-y plane.

10-7 Kepler's law of equal areas

In deriving the fact that the orbit of a planet lies in a single plane, we used only the fact that the *direction* of the planet's angular momentum $\mathbf{L}$ is fixed in time. The magnitude of $\mathbf{L}$ is also constant, and the purpose of this section is to explore the implications of this fact.

Consider, in Figure 10-16, a planet with a trajectory AB lying in the x-y plane at an instant when it has a velocity $\mathbf{v}$. Let us describe the motion of the planet in terms of its polar coordinates (r, θ) as defined in the figure, and let v_r and v_θ represent the components of the velocity $\mathbf{v}$ along and at right angles to the radial direction, respectively. In terms of these quantities the magnitude L of the angular momentum may be found, by use of (10-7), to be

$$L = mrv_\theta \tag{10-11}$$

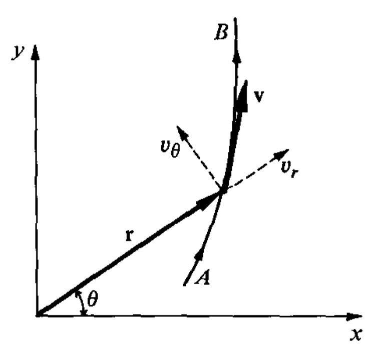


Figure 10-16

since by definition v_θ is the component of **v** perpendicular to **r**. The fact that L is constant thus implies that the product rv_θ is also constant in time. Hence the closer a planet approaches the sun, the larger will be v_θ, but in such a way that the product rv_θ remains constant. Note that (10-11) says nothing whatsoever about the variations in time of the radial component of velocity v_r and the angle θ.

In order to relate the constancy of the magnitude of the angular momentum to Kepler's second law of equal areas, consider in Figure 10-17 a planet that is moving in its orbit about the sun and goes from P_1 to a neighboring point P_2 in an infinitesimal time interval Δt. During this time interval the planet travels a distance $v_\theta \Delta t$ perpendicular to the radius vector, and thereby sweeps out a certain triangular area ΔA. Since the area of a triangle is half the product of its altitude and base, it follows that

$$\Delta A = \frac{1}{2} r v_\theta \Delta t \tag{10-12}$$

Dividing both sides by Δt and taking the limit $\Delta t \to 0$, for the rate dA/dt at which the radius vector sweeps out area, we obtain

$$\frac{dA}{dt} = \lim_{\Delta t \to 0} \frac{\Delta A}{\Delta t}$$
$$= \frac{1}{2} r v_\theta$$

Finally, substituting for the product rv_θ, by use of (10-11) we find that

$$\frac{dA}{dt} = \frac{L}{2m} \tag{10-13}$$

where L is the constant angular momentum of the planet and m its mass. This completes the proof of Kepler's second law.

As an example, suppose in Figure 10-18 that a planet is moving about the sun along the trajectory BC and that t_0 is the time it takes the planet to go from P_1 to P_2. If in the succeeding time interval t_0 the planet goes from P_2 to P_3, then because of (10-13) we may conclude that the area A_1 swept out

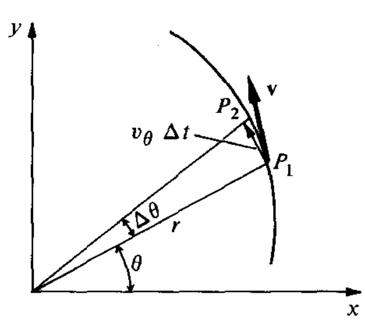


Figure 10-17

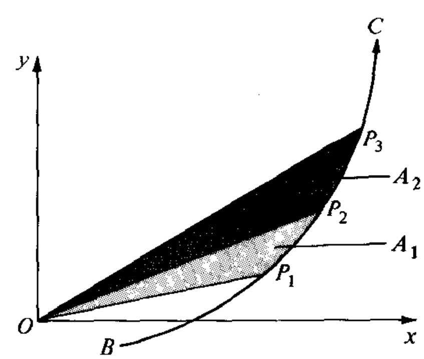


Figure 10-18

during the initial time interval is precisely the same as the area A_2. Moreover, each of these areas A_1 and A_2 is in turn equal to $Lt_0/2m$, where L is the angular momentum of the planet and m is its mass.

Since the period T of a planet is the time it takes it to go once completely around the sun, it follows from (10-13) that

$$T = \frac{2m}{L} A \tag{10-14}$$

where $A \equiv \pi a^2(1-\epsilon^2)^{1/2}$ is the area of the planetary ellipse.

Example 10-6 Calculate the angular momentum L of the planet Jupiter by use of the data in Table 10-1.

Solution Solving (10-14) for the angular momentum L, we obtain

$$L = \frac{2\pi m a^2(1-\epsilon^2)^{1/2}}{T}$$

where all the parameters refer to the orbit of Jupiter. Making use of the data in the table and the fact that the mass of Jupiter is 315 times that of the earth and its semimajor axis is 5.2 times that of the earth, we obtain

$$L = \frac{2\pi \times 315 \times 6.0 \times 10^{24}\ \text{kg} \times (5.2 \times 1.5 \times 10^{11}\ \text{m})^2 \times (1 - 0.05^2)^{1/2}}{3.74 \times 10^8\ \text{s}}$$

$$= 1.9 \times 10^{43}\ \text{J-s}$$

10-8 Circular orbits

Reference to Table 10-1 shows that, with the exception of Mercury and Pluto, the eccentricities of all planets in our solar system are very small. This implies that these planets move about the sun essentially in circular orbits. The purpose of this section is to consider this possibility.

Figure 10-19 shows a planet of mass m moving along a circular orbit of radius a, with the sun at its center. Of the four variables r, θ, v_r, and v_θ defined in Figure 10-16, for circular motion the variable r has the constant value a, and the radial component v_r of its velocity vanishes. Hence, only

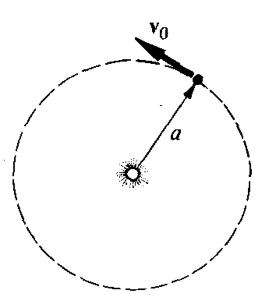


Figure 10-19

the variables θ and v_θ can possibly change as the planet moves about the sun. Moreover, since $r = a$ and the quantity rv_θ must have the constant value L/m, according to (10-11), it follows that v_θ must also be a constant and equal to its initial value v_0. Thus the requirement that the planet orbit about the sun in a circular path implies of necessity that it must do so with a uniform velocity.

Now as the planet moves with uniform velocity in its circular orbit it experiences a force **F** that, according to (10-10), is directed radially inward and has the magnitude GMm/a^2. Equating this to the product of the planet's mass m and its centripetal acceleration, v_0^2/a, we obtain

$$\frac{mv_0^2}{a} = \frac{GMm}{a^2}$$

and this leads to the necessary condition for a circular orbit:

$$v_0^2 a = GM \tag{10-15}$$

If a planet is at a distance a from the sun and is instantaneously traveling tangentially but with a velocity v_0 not satisfying (10-15), then it is not in a circular orbit.

In view of the facts that a circle is an ellipse of zero eccentricity and that Kepler's second law is valid for all central force fields, it follows that Kepler's first two laws are consistent with the circular orbits described by (10-15). To confirm the validity of Kepler's third law for these orbits, we proceed in the following way. The period T of the motion may be expressed as the ratio

$$T = \frac{2\pi a}{v_0}$$

of the distance $2\pi a$ traveled in an orbit and the uniform velocity v_0. Squaring both sides of this relation and substituting for v_0^2 by use of (10-15), we find on dividing by a^3 that

$$\frac{T^2}{a^3} = \frac{4\pi^2}{GM} \tag{10-16}$$

But the right-hand side of this relation is *independent* of all the planetary parameters such as mass, orbital radius, and velocity. Hence the quantity (T^2/a^3) has the *same* value for all planets in the solar system and the validity of Kepler's third law, for the special case of circular orbits, is established. In Section 10-10 it will be shown that (10-16) is also valid for elliptical orbits, provided that the parameter a is interpreted to be the length of the semimajor axis of the planetary ellipse.

10-9 The energy integral

Consider now the more general case, in Figure 10-16, of a planet whose motion is not necessarily circular and for which the specification of all four

variables r, θ, v_r, and v_θ is necessary for a complete description. As the planet moves in orbit about the sun, each of these four variables changes in time, and we shall now describe two important constraints on this variation.

First, because of the constancy of the magnitude of the angular momentum, it follows from (10-11) that the variables r and v_θ are restricted by

$$rv_\theta = \frac{L}{m} \tag{10-17}$$

where the constants L and m are the angular momentum and mass of the planet, respectively. In addition, since the force $\mathbf{F}$ on the planet due to the sun is conservative with a potential energy given in (8-15), it follows that the total energy E of the planet

$$E = \frac{1}{2} m(v_r^2 + v_\theta^2) - \frac{k}{r} \tag{10-18}$$

is also constant in time. In this relation we have used the fact that, according to Figure 10-16, $v^2 = v_r^2 + v_\theta^2$, and have defined the positive constant k by

$$k = GMm \tag{10-19}$$

Taking note of the fact that E and L are for any given planet fixed constants, we see that the relations in (10-17) through (10-19) constitute a complete dynamical description for the motion of the planet. By their use we can predict correctly the elliptical motions of the planets about the sun, as well as the parabolic and hyperbolic trajectories of comets and other celestial objects moving under its influence.

For the special case of circular motion, for which $v_r = 0$, $r = a$, and $v_\theta = \text{constant}$, it is easy to confirm that these relations reduce to the corresponding ones for circular motion in the Section 10-8.

Example 10-7 An α particle (the nucleus of the ^{4}He atom) of mass m is initially moving at a speed v_0 when it is at a very large distance from a fixed gold nucleus, which repels it with a central force varying inversely with the square of the distance r separating them. Suppose, as shown in Figure 10-20, that the α particle is initially moving in such a way that if the force were not acting, its distance of closest approach (the "impact" parameter) to the gold nucleus would be b. Determine the actual distance of closest approach. (*Note*: The process described here is known as

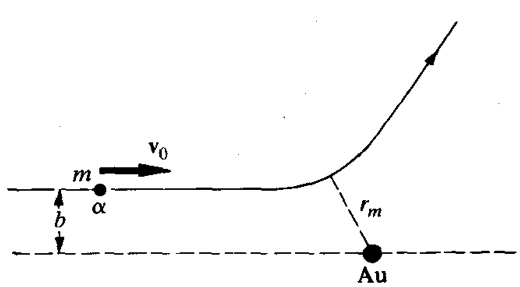


Figure 10-20

Rutherford scattering in honor of Lord Rutherford (1871–1937), who first proposed and carried out studies of the collisions of α particles with various nuclei and thereby gave birth to our present-day picture of the atom; see Section 1-11.)

Solution The potential energy of the α particle is in this case $+k_0/r$, with k_0 a positive constant. Since initially the α particle is very far away from the gold nucleus, its total energy then is all kinetic. Substitution into (10-18) thus yields

$$\frac{1}{2}mv_0^2 = \frac{1}{2}m(v_r^2 + v_\theta^2) + \frac{k_0}{r} \tag{10-20}$$

where v_0 is the initial velocity of the α particle. The initial angular momentum with respect to the position of the gold nucleus of the α particle is mv_0b, so by the conservation of angular momentum we may write

$$mrv_\theta = mv_0b \tag{10-21}$$

Solving (10-21) for v_θ and substituting into (10-20), we find, on dividing throughout by $m/2$, that

$$v_0^2 = v_r^2 + \left\{\frac{v_0b}{r}\right\}^2 + \frac{2k_0}{mr}$$

At closest approach, as shown in Figure 10-20, the α particle is moving perpendicular to the radius vector, so that at this point the radial component v_r of $\mathbf{v}$ vanishes. Substituting $v_r = 0$, we obtain for this distance of closest approach

$$r_m = \frac{k_0}{mv_0^2} + \sqrt{b^2 + \left(\frac{k_0}{mv_0^2}\right)^2}$$

†10-10 Elliptical orbits

The purpose of this section is to generalize the results of Section 10-8 on circular orbits to the case of elliptical ones.

To this end, let us begin with (10-17) and (10-18) and eliminate the variable v_θ between them. The result is

$$E = \frac{1}{2}mv_r^2 + \frac{L^2}{2mr^2} - \frac{k}{r} \tag{10-22}$$

Since $v_r = dr/dt$ and E and L are constants, it follows that the qualitative methods for analyzing bound states discussed in Section 8-7 will be applicable in the present case, provided that the planet is thought of as moving in an effective potential V_e, given by

$$V_e = \frac{L^2}{2mr^2} - \frac{k}{r} \tag{10-23}$$

Figure 10-21 shows a plot of V_e as a function of r. Note that this curve has all the essential features associated with bound-state potentials described in Section 8-7. For example, for very small values of r, the effective potential

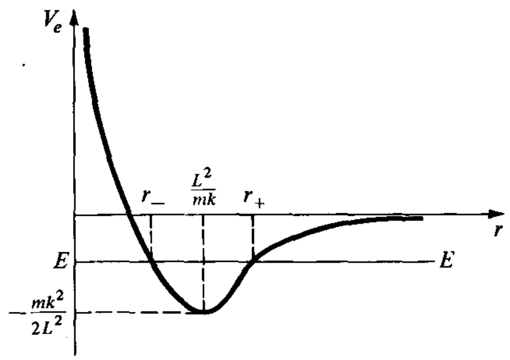


Figure 10-21

V_e is very strong and repulsive; for very large distances, it is very weak and attractive; and for a certain value in between, V_e has a minimum value. Using ideas of elementary calculus, it is easy to show that this minimum occurs at a radial distance L^2/mk and that the minimum value of V_e at this point is $-mk^2/2L^2$.

To proceed, let us assume that, consistent with Kepler's first law, the orbit of a planet is an ellipse and make use of (10-22) to determine its orbital parameters. Let us draw in Figure 10-21 a horizontal line corresponding to a negative value for the energy E, which is, however, greater than the minimum value $-mk^2/2L^2$ of V_e. This horizontal line intersects the V_e-curve at two distinct values for r: call the larger of these r_+ and the other r_-. At these points the radial component of the planet's velocity v_r vanishes. Setting v_r to zero in (10-22), we obtain

$$\frac{L^2}{2mr^2} - \frac{k}{r} - E = 0$$

or, equivalently,

$$r^2 + \frac{kr}{E} - \frac{L^2}{2mE} = 0$$

The roots of this equation are

$$r_\pm = -\frac{k}{2E}\left[1 \pm \left(1 + \frac{2EL^2}{mk^2}\right)^{1/2}\right] \tag{10-24}$$

which, since E is negative but larger than the minimum value of V_e, are both positive and real.

Reference to Figures 10-2 and 10-3 shows that since the sun is located at a focal point of the ellipse, these maximum and minimum values $r_\pm$ are $a(1+\epsilon)$ and $a(1-\epsilon)$, respectively. The root r_+ corresponds to the planet's being at point B in Figure 10-3 and r_- at A. Substitution of these forms for $r_\pm$

into (10-24) leads, after some algebra, to

$$a=-\frac{k}{2E}$$
$$\epsilon=\left[1+\frac{2EL^2}{mk^2}\right]^{1/2} \tag{10-25}$$

or, equivalently,

$$E=-\frac{k}{2a}$$
$$L=[mka\,(1-\epsilon^2)]^{1/2} \tag{10-26}$$

Note that the energy E of the planet depends only on the length of its semimajor axis a; it is independent of the eccentricity ϵ.

The proof of Kepler's third law is now a matter of algebra. Using the fact that the area A of an ellipse is the product of π and the lengths of the semimajor (a) and semiminor ($a\sqrt{1-\epsilon^2}$) axes, from (10-14) we obtain

$$\begin{aligned} T&=\frac{2mA}{L}=\frac{2m\pi a^2(1-\epsilon^2)^{1/2}}{[mka(1-\epsilon^2)]^{1/2}} \\ &=2\pi a^{3/2}\sqrt{\frac{m}{k}} \\ &=\frac{2\pi a^{3/2}}{\sqrt{GM}} \end{aligned} \tag{10-27}$$

The second equality follows by use of (10-26) and the last by use of the definition of k in (10-19). Just as for the case for circular motion, we see that the ratio T^2/a^3 is the constant $4\pi^2/GM$ and is thus independent of all planetary parameters, such as E, L, and m. This then concludes our derivation of Kepler's third law from Newtonian mechanics.

†10-11 Deflection of a falling body on a rotating planet

If a body is dropped from a height h above the surface of the earth, then it is usually thought that it falls vertically downward toward the ground. However, precision measurements show that this is not at all the case, but rather because of the rotation of the earth about its axis there is a very slight eastward deflection of all such falling bodies. The purpose of this section is to analyze this effect.

Physically, this eastward deflection of a falling body may be thought of as being produced by fictitious forces that come into play by virtue of the fact that the earth is an accelerated reference frame. That is, because of the earth's rotation about its axis, each point on its surface undergoes a centripetal acceleration directed radially inward, so that all origins fixed with

respect to the earth's surface are accelerated. The fictitious force associated with this acceleration has been given the name *Coriolis force.*

Equivalently, this deflection can be thought of as arising because of the fact that the trajectory of a particle in the earth's field must be such as to conserve angular momentum. Thus as its radial distance r from the center of the earth becomes smaller, it follows from (10-11) that the particle's tangential component of velocity v_θ must increase.

We shall now adopt this latter point of view and show how this eastward deflection can be explained by making use of the conservation of angular momentum.

Consider, in Figure 10-22, a particle of mass m at a height h above the surface of the earth at a point which we take, for simplicity, to lie on the equator. For cases of physical interest, the height h will usually be very small compared with the radius R_0 of the earth. If the particle is assumed to start out at rest relative to that point on the surface of the earth vertically beneath it, then initially the radial component of the particle's velocity v_r vanishes, and its tangential component v_θ will be $\omega(R_0 + h)$, where ω is the angular velocity of the earth. On being released, because of the gravitational attraction of the earth, the particle starts to fall vertically downward, and thus its radial distance r from the earth's center starts to decrease. It follows from (10-11) that the tangential component of its velocity v_θ must *increase* during this process and in a way such as to keep the product rv_θ constant. In more quantitative terms this means that during its descent to the ground, the radial distance r and the tangential velocity v_θ must be related by

$$mrv_\theta = m\omega\,(R_0 + h)^2 \tag{10-28}$$

since initially the velocity of the particle is $\omega(R_0 + h)$, so that its angular momentum L relative to the center of the earth is $m\omega(R_0 + h)^2$. Anticipating the fact that the eastward deflection will be very small we may write, for the radial distance r of the body from the center of the earth at any time t,

$$r = R_0 + h - \frac{1}{2}gt^2 \tag{10-29}$$

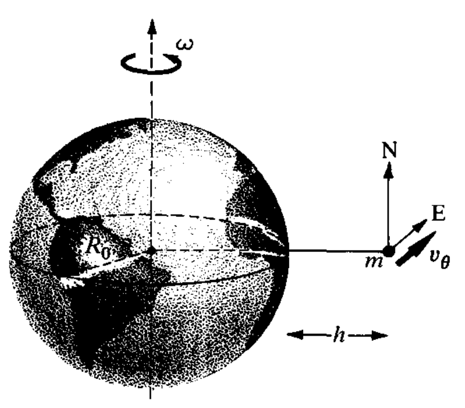


Figure 10-22

and substitution into (10-28) leads to

$$v_\theta = \frac{\omega(R_0+h)^2}{R_0+h-\frac{1}{2}gt^2} = \frac{\omega(R_0+h)}{[1-\frac{1}{2}gt^2/(R_0+h)]} \tag{10-30}$$

To calculate the magnitude of the eastward deflection, let v_y represent, at time t, the velocity of the falling body in the eastward direction as seen by an observer fixed with respect to the surface of the earth. Then

$$\begin{aligned} v_y &= v_\theta - r\omega \\ &= \frac{\omega(R_0+h)}{[1-\frac{1}{2}gt^2/(R_0+h)]} - (R_0+h-\tfrac{1}{2}gt^2)\omega \\ &\cong (R_0+h)\omega\left[1+\frac{\frac{1}{2}gt^2}{R_0+h}\right] - (R_0+h)\omega + \tfrac{1}{2}gt^2\omega \\ &= gt^2\omega \end{aligned}$$

where the second equality follows by use of (10-29) and (10-30) and the third is then obtained by use of the fact that $gt^2 \ll (R_0+h)$ and employing the binomial theorem. Integrating this formula for v_y we obtain, for the total eastward deflection y at time t,

$$y = \left(\frac{1}{3}\right) gt^3\omega$$

Finally, since the time required for the particle to fall the distance h is $(2h/g)^{1/2}$ according to (10-29), it follows that the full eastward deflection D may be expressed as

$$D = \frac{\omega g}{3}\left(\frac{2h}{g}\right)^{3/2} \tag{10-31}$$

If, for example, a particle is dropped from a height of 100 meters, its eastward deflection according to this formula is found (by substituting the values $h = 100$ meters and $\omega = 7.2 \times 10^{-5}$ rad/s) to be 2.2 cm. This deflection is very small and will be observed only under very carefully controlled conditions.

It is important to keep in mind the physical basis for the predicted deflection in (10-31). As the particle descends to the surface of the earth, its tangential velocity v_θ must increase in order to keep the product rv_θ constant. It follows, therefore, that its tangential velocity must exceed that of the point on the surface that initially was immediately below it, and it is therefore deflected towards the east. Using arguments of the same type it is shown in the problems that if a particle is thrown vertically upward from a point on the surface of the earth then when it comes down it will land *west* of the point from which it was originally thrown upward. Why?

Figure 10-23 shows an alternate way of viewing this deflection. As a particle is released at point A, it finds itself under the influence only of the gravitational field of the earth and, in accordance with the analysis of Section 10-10, it will then travel along an elliptical path, such as AB in the

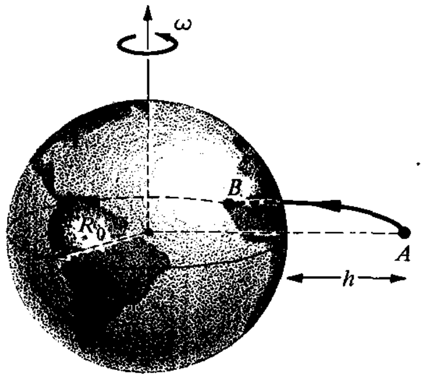


Figure 10-23

figure. However, because $h \ll R_0$, long before it can complete a single orbit, it strikes the earth and at a point given by (10-31).

10-12 Summary of important formulas

If **r** is the position of a particle relative to certain origin O, then its angular momentum **L** relative to this same origin is defined by

$$\mathbf{L} = m\mathbf{r} \times \mathbf{v} \qquad (10\text{-}7)$$

where m is the mass of the particle and **v** its velocity. The torque that a force **F** exerts on a particle is defined by

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \qquad (10\text{-}8)$$

where **r** is the position of the particle relative to the origin O with respect to which $\boldsymbol{\tau}$ is defined. The torque $\boldsymbol{\tau}$ acting on a particle is related to its angular momentum **L** measured with respect to the same *Newtonian origin* by

$$\boldsymbol{\tau} = \frac{d\mathbf{L}}{dt} \qquad (10\text{-}9)$$

The motion of a planet of mass M about the sun is characterized by the two constants of the motion of energy E and angular momentum **L**. In terms of the parameters defined in Figure 10-16, these are

$$L = mrv_\theta \qquad (10\text{-}11)$$

$$E = \frac{1}{2} m (v_r^2 + v_\theta^2) - \frac{k}{r} \qquad (10\text{-}18)$$

with k as defined in (10-19). The rate dA/dt at which the radius vector from the sun to each planet sweeps out area is constant for each planet and is

$$\frac{dA}{dt} = \frac{L}{2m} \qquad (10\text{-}13)$$

with L the angular momentum of the planet and m its mass.

QUESTIONS

1. Define or describe briefly the meaning of the terms (a) torque; (b) right-hand rule; (c) moment arm; and (d) astronomical unit.
2. Is the definition for the magnitude of a cross product, (10-2), consistent with the requirement that the magnitude of all vectors must be nonnegative? Justify your answer.
3. Consider in Figure 10-24 a particle located at a point P on the positive y-axis of a coordinate system. What is the direction of the torque $\boldsymbol{\tau}$ with respect to the origin of this system, produced by a force **F** acting on the particle if the direction of the force is: (a) Along the x-axis? (b) Along the y-axis? (c) Along the z-axis?

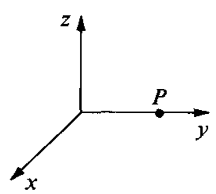


Figure 10-24

4. For the situation in Figure 10-24 is it possible to have a force **F** acting on the particle at P so that the torque it produces about the origin is directed along the y-axis? Explain.
5. What is the direction of the force **F** acting on the particle in Figure 10-24 if the torque $\boldsymbol{\tau}$ is directed: (a) Along the z-axis? (b) Along the x-axis?
6. Review briefly the argument that shows that the orbit of a body moving in the gravitational field of the sun must lie in a single plane. Explain this feature in physical terms.
7. Observations show that, consistent with Kepler's first law, each planet in our solar system travels in an orbit that lies in a single plane. Does the fact that the orbits of different planets lie in different planes contradict Kepler's laws? Explain.
8. The planet Jupiter has 12 moons. Would you expect Kepler's laws to be applicable for the motions of these moons? Why might there be small deviations for some of them?
9. Why is it that the relation in (10-9) is, strictly speaking, *not* valid for an origin at rest with respect to a point on the surface of the earth?
10. Consider, in Figure 10-25, a particle in motion along the inner surface of a *smooth*, hemispherical bowl, so that the only forces acting on it are the downward force of gravity $m\mathbf{g}$ and a normal force **N**. Explain why the component along the vertical of the angular momentum (taken about point C as an origin) is a constant of the motion. If the interior of the bowl were not smooth, would this component of angular momentum still be conserved?

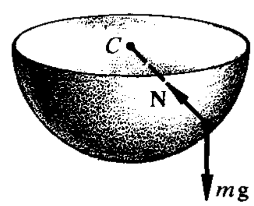


Figure 10-25

11. Consider a pendulum with a very small bob. What is the torque acting on the bob with respect to an origin *at the bob* itself? What is the angular momentum with respect to this origin? Is the relation in (10-9) valid for this case even though this origin accelerates?
12. What are the two points on the surface of the earth at which an object stationary with respect to the surface of the earth will have zero angu-

lar momentum with respect to an origin taken at the earth's center?

13. What is the direction of the angular momentum, with respect to the center of the earth, of an object that is at rest relative to a point on the surface?
14. What must be the magnitude and the direction of the horizontal component of the velocity of a projectile so that its angular momentum with respect to the center of the earth vanishes?
15. Imagine an object that is projected vertically upward at a point on the equator. Explain by use of (10-11) why to an observer stationary with respect to the earth the projectile will not appear to continue to move along the vertical but rather be deflected toward the west.
16. Making use of the effect described in Question 15, explain why it is that if you wish to fire a projectile due north, you must aim the cannon slightly east of north.
17. In view of the fact that our moon is acted upon by the gravitational field of both the earth and the sun, why do we say that the moon moves in a central field? Why do you suppose that, to a very high degree of approximation, the moon's angular momentum about the earth is conserved nevertheless?
18. Precision measurements show that the quantity a^3/T^2 is only approximately constant in the solar system and that, as shown in the fifth column of Table 10-1, small differences between the values of this quantity exist for certain of the planets. Can you think of a simple way to account for such deviations?
19. Describe a way to establish experimentally that the earth rotates about its axis by observing the motion of projectiles near the surface of the earth.
20. Show that for a comet with $E \geqslant 0$, a parabolic or a hyperbolic orbit is consistent with the graph in Figure 10-21. How would you compute the distance of closest approach to the sun of such a comet?

PROBLEMS

1. A particle, which in a certain coordinate system is located at the point $\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$, is acted on by a constant force $\mathbf{F} = \mathbf{i}F_0$. Calculate the torque about the origin produced on the particle by this force.
2. For the particle in Problem 1 calculate the torque $\boldsymbol{\tau}$ that the given constant force produces on the particle, but this time about the point with coordinates $(a, 0, 0)$.
3. A particle moving on a circle of radius 20 cm experiences a torque about the center of the circle of 20 N-m. What is the moment arm? What force $\mathbf{F}$ is required to produce this torque?
4. Suppose that the particle in Figure 10-24 has a mass of 50 grams and is instantaneously located at a distance of 15 cm from the origin. If it has a velocity $\mathbf{v}$ that lies entirely in the y-z plane and has a magnitude of 3 cm/s and makes an angle of 30° with respect to the y-axis, calculate:
   (a) The instantaneous value of its angular momentum about the origin.
   (b) The rate of change of the particle's angular momentum if it is subjected to a force $\mathbf{F}$ with a magnitude of 5.0 newtons and lying in the y-z plane and at an

angle of 45° with respect to the negative sense of the y-axis.

5. Making use of the data in Table 10-1 calculate the angular momentum of Uranus about the sun assuming that it moves with uniform velocity in a circular orbit and that its mass is 14.4 times that of the earth.
6. Assuming that the moon orbits the earth in a circle of radius 3.84×10^5 km with a period of 27.3 days, and that its mass is 0.012 times that of the earth, calculate its angular momentum relative to the center of the earth. Is this its angular momentum relative to the center of the sun? Explain.
7. During the flight of Apollo 11, astronaut Michael Collins circled the moon in an approximately circular orbit in the command module. Assuming that the period of this motion was 90 min and that his orbit was 100 km above the lunar surface, calculate **(a)** his velocity and **(b)** his angular momentum with respect to the center of the moon, assuming that his mass is 80 kg.
8. A small bead of mass m slides with *uniform* speed v_0 on a wire in the form of an ellipse of semimajor axis a and semiminor axis b. Calculate the maximum and minimum values of its angular momentum if **(a)** the origin is taken at the center of ellipse and **(b)** the origin is taken at one of the focal points of the ellipse.
9. A satellite orbits the earth in an elliptical path such that at perigee its distance from the center of the earth is $1.02R_0$, where $R_0 = 6.4 \times 10^3$ km is the radius of the earth, whereas at apogee its separation from the center of the earth is $1.06R_0$. Calculate the length of the semimajor axis of the ellipse and its eccentricity. (*Hint*: See Figures 10-2 and 10-3.)
10. Consider two identical particles, each of mass m and located at the equator. Suppose that one of them is on the surface of the earth and the second is a vertical distance h above the first. Show that if $h \ll R_0$ then the angular momentum of the upper relative to the center of the earth exceeds that of the lower by $2\,mR_0h\omega$, where R_0 is the radius of the earth and ω its angular velocity. What is the direction of their angular momentum?
11. Consider again the physical situation of Problem 10, but suppose this time that the particles are at a latitude $\lambda = 30°$, that each has a mass $m = 1.5$ kg, and that the upper is 100 meters vertically above the lower. Calculate now the difference of their angular momenta relative to the center of the earth.
12. Consider a simple pendulum of length $l = 50$ cm and whose bob has a mass of 50 grams. If it is displaced by an angle 60° and then released at rest, calculate the angular momentum of the bob, relative to the point of suspension, when the bob is at the lowest point of its orbit.
13. Consider a particle of mass m, which is moving with *constant* velocity $\mathbf{v}_0$ with respect to a certain origin. Show that the time rate of change of its angular momentum with respect to that origin is zero. With respect to what other origins is the time rate of change of this angular momentum also zero?

*14. Prove that the relation in (10-9) is invariant under Galilean transformations. That is, show that if (10-9) is valid for a certain origin O, then it is also valid with respect to any other origin O' that moves with uniform velocity $\mathbf{v}_0$ with respect to the first.

(*Hint*: If $\mathbf{r}$ is the instantaneous position of the particle with respect to O, then its position $\mathbf{r}'$ with respect to the origin O' is $(\mathbf{r} - \mathbf{v}_0 t)$.)

15. In a certain coordinate system the coordinates of a particle (in meters) are $(0, 2, -3)$ at an instant when the force $\mathbf{F}$ acting is given in newtons by $\mathbf{F} = 3\mathbf{i} - 2\mathbf{j} + \mathbf{k}$. Find a unit vector along the direction of the torque about the origin produced by the force.

16. Confirm the validity of the distributive property of the cross product in (10-4) for the special case that all three vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ lie in a single plane.

17. Confirm the validity of the relation

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$$

for arbitrary vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.

18. Express the components of the angular momentum L_x, L_y, L_z of a particle of mass m in terms of the components x, y, z of its position and the components mv_x, mv_y, mv_z of its momentum.

19. A small block of mass m starts from rest at the top of a *smooth*, inclined plane of height h and angle α—that is, at point A in Figure 10-26. Let $\mathbf{N}$ represent the normal force acting on the block, and let $\mathbf{v}$ represent its velocity at a subsequent instant when it has moved a distance x down the plane.

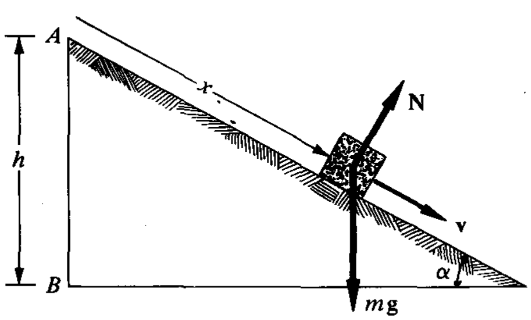


Figure 10-26

(a) Calculate the torque, with respect to an origin at point A, produced by all forces acting on the block.

(b) Calculate the angular momentum $\mathbf{L}$ and its time derivative with respect to the same origin at A.

(c) By comparing the results of (a) and (b) with (10-9) what can you say about the value of the normal force N that the plane exerts on the block?

20. Consider the physical situation in Figure 10-26 of a block of mass m sliding down an inclined plane.

(a) Show that the magnitude of the torque τ (with respect to point B) acting on the block is

$$\tau = mgx \cos \alpha - N(x - h \sin \alpha)$$

(b) What is the angular momentum of the block about point B?

(c) Substitute the answers to (a) and (b) into (10-9) and compare the result with that obtained by direct usage of Newton's second law.

21. Consider the motion of a simple pendulum of mass m and of length l. Show, by an application of (10-9) with the origin at the point of suspension, that the resulting equation may be expressed as

$$\frac{dv}{dt} = -g \sin \theta$$

where v is the tangential component of the velocity of the bob.

***22.** Figure 10-27 shows a "conical" pendulum of length l whose bob is not confined to motion in a single plane. As shown in the figure, let us describe its motion in terms of the angle θ that the string makes with the vertical and in terms of the two components of velocity v_1 and v_2, where v_1 is the component of the

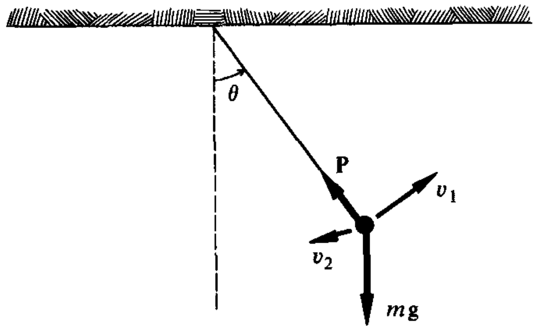


Figure 10-27

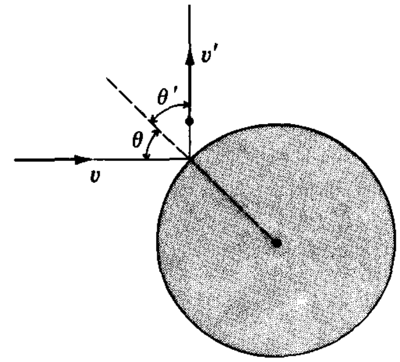


Figure 10-28

velocity of the bob in the plane determined by the string and **g**, and v_2 is the component of its velocity at right angles to this plane.

(a) Explain why the component along the vertical direction of the torque about the point of suspension vanishes.

(b) Show that the quantity $mbv_2 \sin\theta$ must be a constant of the motion.

(c) Explain why the quantity E, given by

$$E = \tfrac{1}{2}m(v_1^2 + v_2^2) + mgl(1 - \cos\theta)$$

is also a constant of the motion in this case.

23. A conical pendulum 1.5 meters in length is given an initial displacement of 30° from the vertical and an initial velocity of 0.5 m/s in a direction perpendicular to the plane determined by the string and the direction of gravity. Make use of the results of Problem 22 to determine the values for the velocity components v_1 and v_2 at the subsequent instant when the bob is at an angle of 20° with respect to the vertical.

24. A particle of mass m approaches with a speed v a spherical object of radius a and strikes it at an angle θ, as shown in the Figure 10-28. Assume that the force the sphere exerts on the particle when they are in contact is directed along a radius.

(a) Show that

$$v \sin\theta = v' \sin\theta'$$

(*Hint*: Why is the angular momentum of the particle conserved?)

(b) Under what circumstances will the angle of incidence θ be the same as the angle of reflection θ'?

25. Consider a bead of mass m moving with uniform velocity v_0 around a circle of radius a. Show that (10-14), which relates the period of the motion and the area of the circle, is an identity for this case; that is, evaluate each of the quantities T, A, and L in terms of the given parameters m, v_0, and a and thus show that (10-14) is satisfied as an identity.

26. An electron of mass $m = 9.1 \times 10^{-31}$ kg orbits a stationary proton in a circle of radius $r_0 = 0.53 \times 10^{-10}$ meter in a hydrogen atom. According to classical ideas, the electron remains in the circular orbit because of the electrical attraction between it and the proton corresponding to an attractive central force of strength k/r^2, where k has the value $k = 2.3 \times 10^{-28}$ N-m².

(a) Determine the velocity of the electron.

(b) What is the angular momentum L of the electron?

(c) Calculate the total energy E of the electron.

***27.** A particle of mass m is moving in Figure 10-20 with a speed v_0 when it is at a very large distance from a repulsive scattering center corresponding to the central force $\mathbf{F} = (k/r^4)\mathbf{r}$, where k is a positive constant. Suppose that the particle is moving in such a way that if the force were not acting, its distance of closest approach would be b.

(a) What is the potential energy of the particle when it is at a distance r from the force center?

(b) Write down two relations involving the components v_r and v_θ of the velocity and the given parameters k, m, and b.

(c) Show that the distance of closest approach r_m is

$$r_m = b\left[1 + \frac{k}{mv_0^2 b^2}\right]^{1/2}$$

(d) What is the velocity of the particle at the instant when its separation distance from the force center is a minimum?

***28.** Suppose a body of mass 2 kg orbits about a force center corresponding to a potential energy

$$V(r) = -\frac{1}{r}$$

where V is in joules and r is in meters. Assume that the particle has an orbital angular momentum $L = 1$ J-s.

(a) Write down two relations based on the conservation laws of energy and of angular momentum.

(b) Show that for any negative value of the total energy E that is greater than -1 joule there exists a bound state.

(c) Show that the largest and smallest values r_+ and r_-, respectively, for the separation of the body from the force center are given by

$$r_\pm = -\frac{1}{2E} \mp \frac{1}{2E}(1+E)^{1/2}$$

29. Consider again the motion of the electron in Problem 26 in a potential energy field $-k/r$, where $k = 2.3 \times 10^{-28}$ N-m^2. Niels Bohr in 1911 developed a model for the hydrogen atom by postulating that the angular momentum L of an electron in a hydrogen atom is not arbitrary but can only assume the values given by

$$L = n\hbar \qquad n = 1, 2, 3, \ldots$$

where $\hbar$ is $(2\pi)^{-1}$ times Planck's constant and has the value $\hbar = 1.06 \times 10^{-34}$ J-s. **(a)** Show that for this model the total energy E_n of the electron is restricted to the values

$$E_n = -\frac{k}{2a_0 n^2} \qquad n = 1, 2, 3, \ldots$$

where $a_0 = \hbar^2/km$. **(b)** Obtain a numerical value for a_0 and for E_1 and E_2.

30. A particle of mass m moves with uniform velocity v_0 through free space just prior, as shown in Figure 10-29, to its entering a circular region of radius a where there is an attractive potential of strength $-V_0$

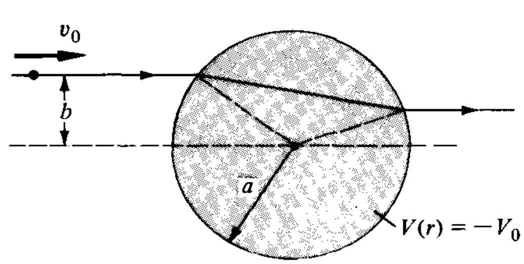


Figure 10-29

($V_0 > 0$). Assume that as shown in the figure the impact parameter b is less than the radius a.

(a) Write down the values of the energy E and of the angular momentum L in terms of m, v_0, and b.

(b) What is the kinetic energy of the particle when it is inside the region $r < a$? What is its velocity in this region?

(c) Show that the distance of closest approach to the scattering center is $b/[1 + 2V_0/mv_0^2]^{1/2}$.

31. By differentiating V_e in (10-23), show that its minimum value occurs at a radial distance L^2/mk and that this minimum value is as given in Figure 10-21. Show also that at this minimum value of r, call it a, the relation in (10-15) is satisfied.

†32. Making use of (10-27), calculate the period of the satellite in Problem 9.

†33. Using the value for the solar mass of 2.0×10^{30} kg, calculate the value of the constant a^3/T^2 in the solar system by solving for this quantity by use of (10-27). Compare your result with that in the fifth column of Table 10-1.

†34. Using the values for a and ϵ from Table 10-1, calculate the values for E and L for the earth by use of (10-26). Compare with the answers you would have gotten for E and L assuming the earth to orbit about the sun in a circle of radius 1.5×10^8 km.

†35. Repeat Problem 34, but this time calculate E and L for the planet Mercury. Compare your results with those that would be obtained if Mercury moved in a circle of radius numerically equal to the length of its semimajor axis.

†36. A satellite of mass m orbits a planet in a circle of radius a and velocity v_0 related in accordance with (10-15). Suppose that it suddenly acquires a *radial* component of velocity numerically equal to $v_0/2$.

(a) What is the value of its angular momentum in terms of m, a, and v_0?

(b) Show that the new value for its total energy E is three quarters of its original value.

(c) Making use of (10-26), calculate the semimajor axis and the eccentricity of the ellipse as a result of the change in velocity.

†37. A satellite orbits a planet in a circular orbit of radius a and with a velocity v_0. Suppose it is suddenly struck in such a way that its velocity perpendicular to the radius vector is increased from its original value v_0 to the value $2v_0$.

(a) What is the new value for the angular momentum of the satellite in terms of the parameters m, a, and v_0?

(b) Calculate the change in the total energy E of the satellite and determine whether or not the satellite is still in a bounded orbit.

†38. A particle located at the equator is thrown vertically upward from the ground with an initial velocity v_0.

(a) Is its angular velocity (relative to a system with origin at the center of the earth but *not* rotating with it) larger or smaller when in flight compared with its angular velocity when on the surface of the earth? (*Hint*: The particle's angular momentum must be conserved.)

(b) Based on your result to (a), will the particle be deviated east or west of its point of origin when it returns to the ground?

(c) By analogy to the technique used to derive (10-31), calculate

the magnitude of this deviation D and show that it is equal to

$$D = \frac{2}{3}\frac{v_0^3}{g^2}\omega$$

†39. Calculate the magnitude of the deflection D in Problem 38 for the case in which $v_0 = 50$ m/s.

†40. A particle is dropped from a height h above the surface of the earth at a latitude λ in the northern hemisphere. Show that it is deflected eastward by an amount

$$\frac{\cos\lambda}{3}\omega g\left(\frac{2h}{g}\right)^{3/2}$$

What would be the direction and magnitude of this deflection if it were dropped at a latitude λ in the *southern* hemisphere?

†41. Repeat Problem 40, but assume this time that the particle is thrown vertically upward from the ground at a latitude λ with an initial speed v_0.

11 The dynamics of systems of particles

What is now proved, was once only imagined.
WILLIAM BLAKE

11-1 Introduction

Consider, in Figure 11-1, a rigid block sliding down an inclined plane. Since each element of the block has the same acceleration **a**, there is no ambiguity in our applying Newton's second law

$$\mathbf{F} = m\mathbf{a} \tag{11-1}$$

to analyze the motion of the block. Indeed, this is precisely what we have done in the past, although, strictly speaking, (11-1) is applicable only to point particles.

In addition to simple macroscopic systems of this type there are other ones for which different parts of the system may have varying accelerations and for which, therefore, (11-1) is not directly applicable. Consider, for example, in Figure 11-2, a wheel rolling at a uniform velocity $\mathbf{v}_0$ along a horizontal and rough surface. The center A of the wheel moves at the uniform speed v_0 and thus is not accelerating. On the other hand, the point B, which is instantaneously the top point of the wheel, is also moving in a

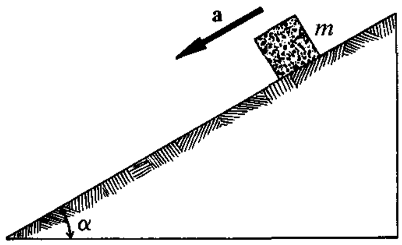


Figure 11-1

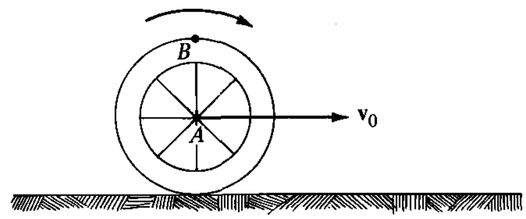


Figure 11-2

vertical circle and thus has a centripetal acceleration directed vertically downward. In general, different points of the rolling wheel have different accelerations and thus we cannot apply (11-1) to this case. Similarly, since the tip of the propeller blade of an airplane in flight has a different acceleration than does the body of the plane itself, we cannot apply (11-1) directly to a plane either.

The basic problem confronting us in a consideration of the dynamics of multicomponent systems therefore has two facets: (1) the need to justify the usage of (11-1) for a system such as the one in Figure 11-1, each of whose components has the same acceleration; and (2) the need to find a means for describing the dynamics of systems whose constituents do not all have the same acceleration. The purpose of this chapter is to study this problem and to develop in this connection the two basic theorems that govern the dynamics of these complex systems.

To derive these theorems we shall make the basic assumption that all macroscopic systems consist of a very large number of point particles which may be thought to be its constituent atoms and molecules. Moreover, we shall suppose that the dynamics of each of these constituents is governed by (11-1), with **F** the total force on each particle. In general, **F** will consist of the sum of two types of forces. These are (1) *external forces*, such as gravitational or electromagnetic forces; and (2) *internal forces*, such as interatomic or intermolecular forces which the microscopic constituents of the system exert on each other. In principle, given the external and internal forces acting on each particle, the dynamics of the system as a whole is governed by the collection of the Newtonian equations of motion for each particle. Therefore, these relations will form the basis for our analysis of macroscopic systems.

11-2 Center-of-mass motion of a two-body system

To help fix ideas, let us consider first the case of a two-body system in an external force field. Although, strictly speaking, the appropriate term here should be "particles" and not "bodies," anticipating the theorem established in Section 11-4 we shall frame our discussion in terms of bodies.

Consider, in Figure 11-3, two bodies of masses m_1 and m_2 moving at the

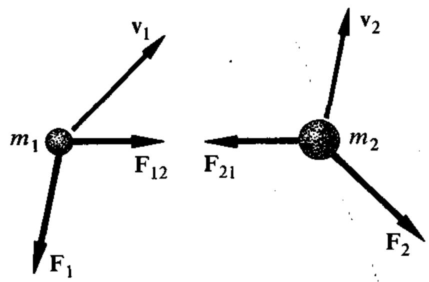


Figure 11-3

instantaneous velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively. Let $\mathbf{F}_1$ and $\mathbf{F}_2$ represent any external forces, such as that of gravity, acting on m_1 and m_2, respectively. In addition to these external forces, each body will also, in general, experience a force due to the proximity of the other, and let us use the symbol $\mathbf{F}_{12}$ to represent this *internal* force on m_1 due to m_2. Correspondingly, let $\mathbf{F}_{21}$ be the force on m_2 due to m_1. In accordance with Newton's second law, m_1 accelerates under the total force $(\mathbf{F}_1+\mathbf{F}_{12})$ acting on it, and similarly m_2 moves under the action of a total force $(\mathbf{F}_2+\mathbf{F}_{21})$. Although it has been assumed in the figure that the internal forces $\mathbf{F}_{12}$ and $\mathbf{F}_{21}$ lie along the line joining the bodies, this need not be the case. However, according to Newton's third law they must be equal and opposite:

$$\mathbf{F}_{12} = -\mathbf{F}_{21} \tag{11-2}$$

Now although the individual motions of the two bodies in Figure 11-3 may be complex, the dynamical behavior of their center of mass is independent of the internal forces and is consequently much simpler. The specific purpose of this section is to establish the following basic result:

The product of the total mass of the system and the acceleration of the center of mass is equal to the vector sum of the external forces acting on the two bodies:

$$\mathbf{F}_1+\mathbf{F}_2 = M\frac{d\mathbf{V}_c}{dt} \tag{11-3}$$

with $M = m_1+m_2$ the total mass of the system and with $\mathbf{V}_c$ the velocity of the mass center

$$\mathbf{V}_c = \frac{1}{M}(m_1\mathbf{v}_1+m_2\mathbf{v}_2) \tag{11-4}$$

which is defined as the time derivative of the center-of-mass position vector

$$\mathbf{R}_c = \frac{1}{M}(m_1\mathbf{r}_1+m_2\mathbf{r}_2) \tag{11-5}$$

In other words, the motion of that fictitious point which lies along the line joining two bodies, and called their center of mass, moves as if it were a particle of mass equal to the total mass $M = m_1 + m_2$ of the system and were acted upon by the sum of the external forces (only) acting on it (see Figure 11-4).

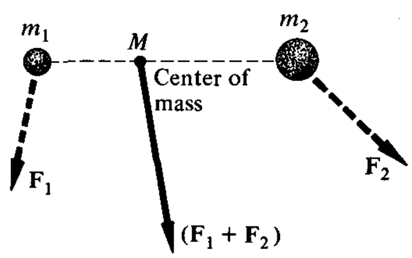


Figure 11-4

Note the very important feature that as far as the motion of the mass center is concerned, the internal forces $\mathbf{F}_{12}$ and $\mathbf{F}_{21}$ play absolutely no role. This is very important, since the external forces acting on a system are generally much better understood, and thus easier to deal with, than the internal ones. For the special case of an isolated system for which the sum of the external forces $(\mathbf{F}_1 + \mathbf{F}_2)$ vanishes, (11-3) reduces, as it must, to the conservation-of-momentum law considered in the previous chapter. That is, if the sum of the external forces acting on the two-body system vanishes, then, consistent with (11-3), the center of mass moves with uniform velocity in a straight line.

The proof of this very important result in (11-3) is surprisingly simple. Since the dynamics of each of the bodies in Figure 11-3 is governed by Newton's second law, it follows that

$$
\begin{aligned}
m_1\frac{d\mathbf{v}_1}{dt} &= \mathbf{F}_1 + \mathbf{F}_{12} \\
m_2\frac{d\mathbf{v}_2}{dt} &= \mathbf{F}_2 + \mathbf{F}_{21}
\end{aligned}
\tag{11-6}
$$

Adding these equations, we find that the internal forces appear only in the combination $(\mathbf{F}_{12} + \mathbf{F}_{21})$, which, according to (11-2), vanishes. Hence the result of this addition is

$$
\begin{aligned}
\mathbf{F}_1 + \mathbf{F}_2 &= m_1\frac{d\mathbf{v}_1}{dt} + m_2\frac{d\mathbf{v}_2}{dt} \\
&= M\frac{d\mathbf{V}_c}{dt}
\end{aligned}
$$

where the last equality follows from the definition of $\mathbf{V}_c$. The basic result in (11-3) is thereby established.

Example 11-1 Two bodies of respective masses m_1 and m_2 are connected by a spring and lie on a horizontal and smooth surface as shown in Figure 11-5. Suppose the force $\mathbf{F}_0$ pulls m_2 to the right.

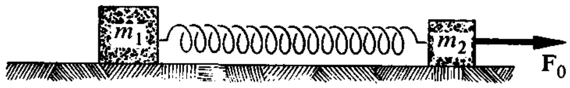


Figure 11-5

(a) What is the acceleration of the center of mass of the system?

(b) If the spring snaps, what is now the acceleration of the center of mass, assuming that $\mathbf{F}_0$ still pulls on m_2?

(c) What are the accelerations of m_1 and of m_2 after the spring has snapped?

Solution

(a) Since the sum of the *external* forces acting on the two bodies is $\mathbf{F}_0$ and since the sum of the masses is $(m_1 + m_2)$, it follows from (11-3) that the acceleration $\mathbf{a}_c$ of the center of mass of the system is

$$\mathbf{a}_c = \frac{\mathbf{F}_0}{m_1 + m_2}$$

(b) The acceleration of the mass center of any system is independent of the internal forces acting between its constituents. Hence the acceleration of the center of mass is the same, regardless of whether or not the spring is attached to the two bodies. Accordingly, $\mathbf{a}_c$ has the same value here as in (a).

(c) The total force acting on m_2 is now $\mathbf{F}_0$. Hence its acceleration $\mathbf{a}_2$ is given by

$$\mathbf{a}_2 = \frac{1}{m_2}\mathbf{F}_0$$

The acceleration $\mathbf{a}_1$ of m_1 vanishes since there are no unbalanced forces acting on it.

Example 11-2 Figure 11-6a shows two blocks of masses m_1 and m_2 connected by a spring of constant k and of natural length l and moving at a uniform velocity v_0 on a smooth, horizontal surface. Suppose that m_1 strikes a vertical wall and comes to rest.

(a) What is the velocity of the center of mass of the system prior to the collision?

(b) What is the velocity of the center of mass immediately after the collision?

(c) What is the acceleration of the center of mass at a subsequent instant when, as shown in Figure 11-6b, the spring has an extention d and m_2 is moving to the left?

Solution Since all motion is confined to one dimension, only the components of velocity and acceleration along this direction are of interest. Let us assume that the positive sense of a coordinate axis is directed to the right in the figure.

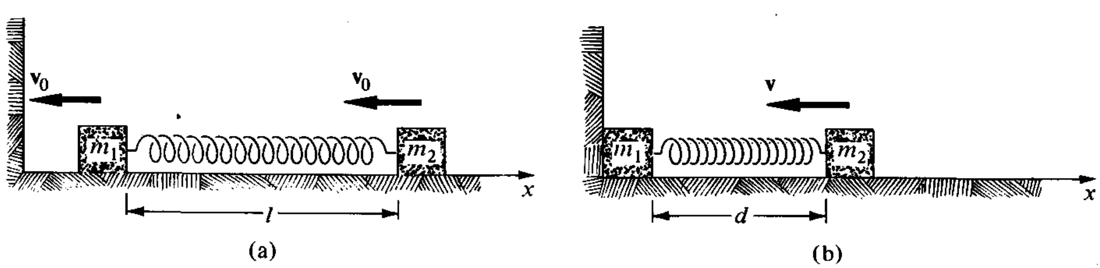


Figure 11-6

(a) Since both bodies travel at the velocity $\mathbf{v}_0$, it follows that

$$V_c = -v_0$$

where the minus sign follows from the fact that the motion is directed toward the left.

(b) Immediately after m_1 has come to rest, m_2 continues to move at the velocity $\mathbf{v}_0$, directed to the left. Substituting the values $v_1 = 0$ and $v_2 = -v_0$ into the appropriate component of (11-4), we obtain

$$V_c = \frac{1}{M}(0 - m_2 v_0)$$

$$= -\frac{m_2 v_0}{m_1 + m_2}$$

(c) Since m_1 is not accelerating, and since for the situation shown in Figure 11-6b the spring exerts on m_1 a force of magnitude $k(l-d)$, directed to the left, it follows that the wall must exert an equal force, but directed to the right. Thus the net *external* force acting on this two-body system is a force of strength $k(l-d)$ directed to the right. Substituting this force into (11-3), we find for the acceleration a_c of the mass center

$$a_c = \frac{k(l-d)}{m_1 + m_2}$$

Example 11-3 A projectile is fired from the ground with a certain muzzle velocity $\mathbf{v}_0$ at an angle of elevation α. Suppose that at the height of its trajectory it explodes into two equal fragments, one of which, call it A, drops from rest vertically downward to the ground. What is the trajectory of the other fragment, B?

Solution According to (11-3), the center of mass of the shell must continue to travel along its initial parabolic trajectory (see Figure 3-27). With this in mind, the solution is shown in Figure 11-7. The dashed line represents the trajectory of the projectile had it not exploded, and this then also represents the motion of its center of mass. Since projectile A falls vertically downward, and since the fragments have equal mass, at any point along its path, the center of mass must lie midway between the fragments. If, for example, fragment A is at position A_1, and B is at B_1, then the location of the center of mass C_1 at this instant must bisect the line A_1B_1. In particular, then, fragment B will strike the ground at a point corresponding to a range $(3v_0^2 \sin 2\alpha)/2g$ and at the same time that fragment A strikes the ground vertically beneath the point of explosion.

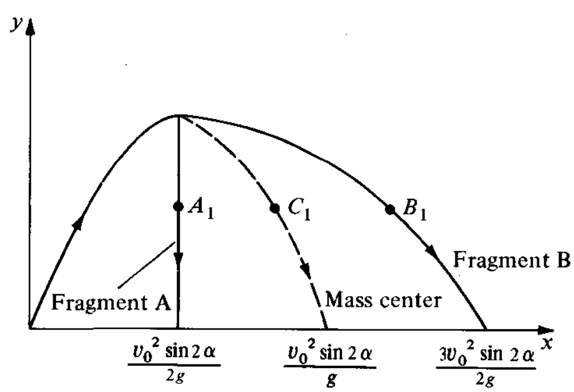


Figure 11-7

11-3 The center of mass of a multiparticle system

The theorem in (11-3) on the motion of the mass center or the center of mass of two bodies is easily generalized to other physical systems. In preparation for this more general discussion, in the present section the center of mass of a composite system will be defined.

Consider, in Figure 11-8, N particles of respective masses $m_1, m_2, \ldots, m_N$, and let their positions relative to a fixed origin O be given by the position vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N$. The total mass M of this system is

$$\begin{aligned} M &= m_1 + m_2 + \cdots + m_N \\ &= \sum m_i \end{aligned} \tag{11-7}$$

where the sum is to be carried out over all particles in the collection. We define the center of mass $\mathbf{R}_c$ of this collection of particles by

$$\begin{aligned} \mathbf{R}_c &= \frac{1}{M}(m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + \cdots + m_N\mathbf{r}_N) \\ &= \frac{\Sigma m_i \mathbf{r}_i}{\Sigma m_i} \end{aligned} \tag{11-8}$$

where again the sums are to be carried out over all particles in the system. For the special case $N = 2$, this is equivalent to (11-5). Qualitatively speaking, $\mathbf{R}_c$ represents a geometric point that is located at the "average" position of the particles weighted in proportion to their masses. If the particles all have the same mass, it follows from (11-8) that $\mathbf{R}_c$ coincides with the geometric center of the system.

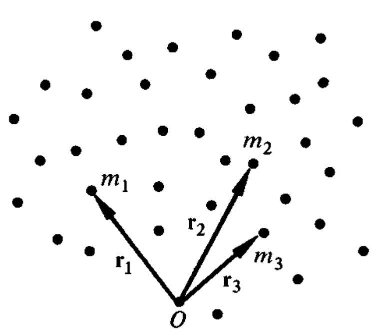


Figure 11-8

In actual calculations of the center of mass, the components of (11-8) are often most useful. In terms of X_c, Y_c, and Z_c, the components of $\mathbf{R}_c$ in a Cartesian coordinate system, (11-8) is equivalent to

$$X_c = \frac{\Sigma m_i x_i}{\Sigma m_i} \qquad Y_c = \frac{\Sigma m_i y_i}{\Sigma m_i} \qquad Z_c = \frac{\Sigma m_i z_i}{\Sigma m_i} \tag{11-9}$$

where x_i, y_i, and z_i are the coordinates of the ith particle in the given coordinate system.

The extension of the definition in (11-9) to continuous distributions of matter is also of practical interest. Consider in Figure 11-9 a solid body and let Δm be an arbitrary infinitesimal mass element of this body. For this case, the first equation in (11-9) becomes

$$\bar{X}_c = \frac{\Sigma x \, \Delta m}{\Sigma \, \Delta m}$$

where x is the x-component of the mass element Δm and the sum is to be carried out over all mass elements in the body. Similar relations apply for the y- and z-components. The bar over $\bar{X}_c$ serves as a reminder that the quantities defined in this way depend in general on the choice made for the mass elements Δm. The coordinate of the mass center is defined to be the limit of these sums as the mass elements tend to zero. Making use of the definition of a definite integral, we have then

$$X_c = \frac{\int x \, dm}{\int dm} \qquad Y_c = \frac{\int y \, dm}{\int dm} \qquad Z_c = \frac{\int z \, dm}{\int dm} \tag{11-10}$$

where the integrals are to be carried out over the entire mass of the body. In a similar way the analogue of (11-7) for continuous distributions of matter is found to be

$$M = \int dm \tag{11-11}$$

Although, in principle, the center of mass of any continuous distribution of matter may be calculated by use of these formulas, sometimes it is easier to proceed by using arguments of symmetry. For example, the center of mass of a homogeneous body—that is, a body having a uniform distribution of mass—must coincide with the geometric center of the body. Thus the center of mass of a homogeneous sphere is at its center, the center of mass of a homogeneous rod is at its midpoint, and the center of mass of a plane lamina in the shape of an equilateral triangle is at the intersection of its three angle bisectors. More generally, if a homogeneous body has any line of symmetry, then the center of mass must lie somewhere along this line. For example, the center of mass of a homogeneous cone must lie along its axis.

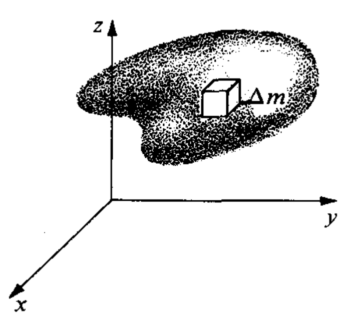


Figure 11-9

Example 11-4 Four particles of masses 1 kg, 2 kg, 3 kg, and 4 kg, respectively, are at the vertices of a rectangle of sides a and b (see Figure 11-10). Assuming that $a = 1$ meter and $b = 2$ meters, determine the location of the center of mass.

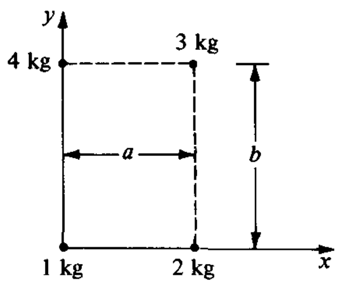


Figure 11-10

Solution Let us set up a Cartesian coordinate system, with the origin at the 1-kg particle and so that all the particles lie in the x-y plane. The coordinates of the four particles are, in increasing orders of their mass $(0, 0)$; $(a, 0)$; (a, b); and $(0, b)$. The total mass M is

$$M = m_1 + m_2 + m_3 + m_4 = 10 \text{ kg}$$

Substituting the given data into (11-9), we find that

$$X_c = \frac{1}{10 \text{ kg}}(0 + a\ 2 \text{ kg} + a\ 3 \text{ kg} + 0) = \frac{a}{2} = 0.5 \text{ m}$$

$$Y_c = \frac{1}{10 \text{ kg}}(0 + 0 + b\ 3 \text{ kg} + b\ 4 \text{ kg}) = 0.7b = 1.4 \text{ m}$$

Example 11-5 Determine the center of mass of a homogeneous rod of length l.

Solution Let us set up a coordinate system as in Figure 11-11, with the origin at one end of the rod and with the rod along the x-axis. Consider the rod as a collection of small mass elements and let a typical one of these at a distance x from the origin have a length dx. In terms of the mass per unit length λ of the rod, the mass dm of this element is

$$dm = \lambda\, dx$$

Substituting into (11-11) and noting that x ranges over all values from 0 to l we find,

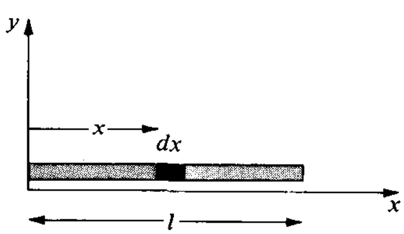


Figure 11-11

for the total mass M,

$$M = \int dm = \int_0^l \lambda\, dx$$
$$= \lambda \int_0^l dx = \lambda l$$

where the third equality follows since the rod is homogeneous and therefore λ is a constant.

In the same way, substituting into (11-10) we find that the coordinate X_c of the center of mass is

$$X_c = \frac{1}{M}\int x\, dm = \frac{1}{\lambda l}\int_0^l x\lambda\, dx$$
$$= \frac{1}{\lambda l}\lambda \int_0^l x\, dx = \frac{1}{l}\frac{x^2}{2}\bigg|_0^l$$
$$= \frac{l}{2}$$

Thus, as we might anticipate based on arguments of symmetry, the mass center is at the center of the rod.

11-4 Dynamics of the center of mass—Theorem I

Consider a system of N particles of respective masses $m_1, m_2, \ldots, m_N$. Let $\mathbf{F}_1$ be the external force acting on m_1, $\mathbf{F}_2$ the corresponding external force on m_2, and so forth. The *total* external force acting on the entire system is thus $\Sigma \mathbf{F}_i$, with the sum carried out over all members of the collection. In addition to these forces, each particle will, in general, experience a force due to the other particles; let us represent these by the symbols $\mathbf{F}_{12}$, $\mathbf{F}_{34}$, and so forth, where $\mathbf{F}_{ij}$ represents the force on the ith particle due to the presence of the jth. Thus $\mathbf{F}_{12}$ is the force on m_1 due to m_2, and similarly $\mathbf{F}_{43}$ is the force on particle 4 due to particle 3. Figure 11-12 shows the situation for the special

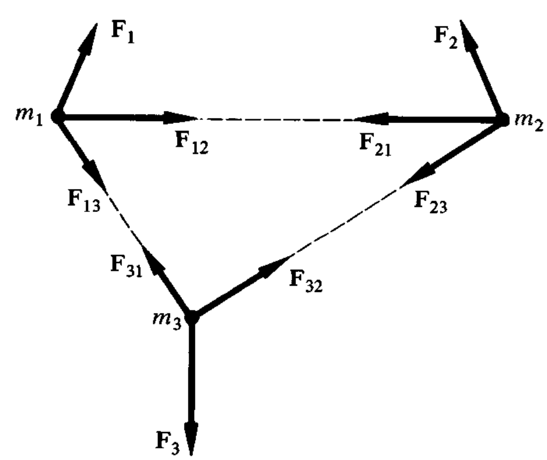


Figure 11-12

case of three particles. Note that the internal forces acting between each pair of particles have been assumed to lie along the line adjoining them. In addition, these internal forces must be consistent with Newton's third law so that we have

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} \tag{11-12}$$

An important characterization of this system is provided by the following theorem:

Theorem I

The product of the total mass of a collection of particles and the acceleration of its center of mass is equal to the vector sum of the *external* forces acting on the particles. In mathematical terms,

$$\mathbf{F} = M\frac{d\mathbf{V}_c}{dt} \tag{11-13}$$

where $\mathbf{F} \equiv \Sigma\mathbf{F}_i$ is the sum of the *external* forces acting on the system and where $\mathbf{V}_c$ is the velocity of the center of mass, which may be obtained, by differentiating (11-8), as

$$\mathbf{V}_c = \frac{1}{M}\sum_i m_i\mathbf{v}_i \tag{11-14}$$

The proof of this theorem is a straightforward extension of the corresponding two-particle proof in Section 11-2. Details will be found in Appendix D.

It is important to note that the force $\mathbf{F}$ that appears on the left-hand side in (11-13) is that produced by *external* forces exclusively. The forces the particles exert on each other play absolutely no role as far as the dynamics of the center of mass is concerned. For the special case that the physical system of interest is the entire universe—that is, all of the planets, stars, galaxies, cosmic dust, and so forth—there are presumably no agents left anywhere to produce external forces, and in this case all forces acting on this system must be internal. Assuming therefore that these internal forces satisfy Newton's third law, we find by applying Theorem I that the acceleration of the center of mass of the entire universe vanishes; or, equivalently, we may say that the *center of mass of the universe moves with uniform velocity in a straight line.*

With Theorem I available, we are able to justify our previous treatment of macroscopic bodies, such as rods, blocks, and planets, as particles, For if all points of a body have the same acceleration $\mathbf{a}$ it then follows that so does its center of mass, and (11-13) is the same as Newton's law for a particle. More generally, Theorem I shows that as far as the motion of the center of mass of a system is concerned, it can be replaced by a single particle—of mass equal to the sum of the masses of its constituents—located at the center of mass and with all external forces acting on it. Consider, for example, a rod of total mass

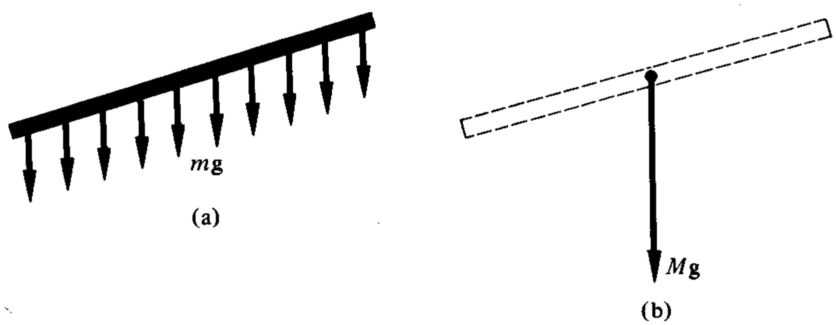


Figure 11-13

M in the uniform gravitational field of the earth, as shown in Figure 11-13. If the rod is thought of as consisting of a large collection of particles each of mass m, then, as shown schematically in Figure 11-13a, each of these particles experiences a downward force $m\mathbf{g}$. Theorem I then compels the conclusion that as far as the motion of the mass center of the rod is concerned, it can be replaced, as in Figure 11-13b, by a single particle of mass M equal to that of the rod and acted upon by a force equal to the vector sum of the gravitational forces acting on the individual particles.

Example 11-6 Four particles, each of mass m, are connected by six identical springs and joined together to form an equilateral tetrahedron, as shown in Figure 11-14. Suppose a constant force $\mathbf{F}_0$ is applied to one of these particles. Find the displacement at time t of the center of mass of this system.

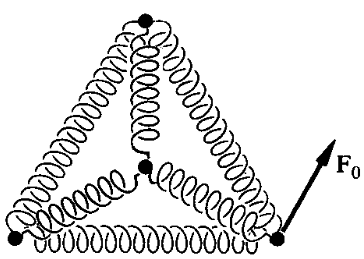


Figure 11-14

Solution Since there are four particles, each of mass m, and since the only external force acting on the system is the constant force $\mathbf{F}_0$, we find by use of (11-13) that

$$4m\frac{d\mathbf{V}_c}{dt} = \mathbf{F}_0$$

Since $\mathbf{F}_0$ is constant, it follows, assuming that initially $\mathbf{V}_c$ vanishes, that this integrates to

$$4m\mathbf{V}_c = \mathbf{F}_0 t$$

Finally, assuming that initially the center of mass of the system is at the origin, by a second integration we find the location of the center of mass at time t to be

$$\mathbf{R}_c(t) = \frac{1}{8m}\mathbf{F}_0 t^2$$

11-5 Energy conservation

In our studies of the law of energy conservation for a single particle we found that, for conservative forces, the sum of the kinetic and the potential energy of the particle is a constant of the motion. The purpose of this section is to describe the circumstances for which there exists an analogous energy conservation law for a multicomponent system.

First, the formal analogy between (11-13) and Newton's second law for a single particle in (4-7) suggests that if the total external force acting on a many-particle system is conservative, then there should be a law of energy conservation associated with its center-of-mass motion. That this is indeed the case will be established in the problems. It should be emphasized, however, that this conservation law deals only with the velocity and position of the *center of mass* of the system; it is not a conservation law associated with any other possible motions of the system.

In order to explore the possibility of the conservation of energy associated with motion other than that of the mass center, let us consider first the case of an isolated many-particle system. For this system, we saw in Chapter 10 that the total momentum $\Sigma\, m_i\mathbf{v}_i$ is a conserved quantity. In addition, it will be established in the problems that, provided the internal forces between particles are conservative, the total energy of the system is also conserved. To illustrate the nature of this conservation law, consider the case of a two-particle system and suppose that the force $\mathbf{F}_{12}$ between the particles depends only on the distance $|\mathbf{r}_1 - \mathbf{r}_2|$ between them and is derivable from a potential energy function V. Then it is shown in Problem 20 that the total energy E of this isolated two-particle system defined by

$$E = T + V \tag{11-15}$$

is a constant of the motion. In this relation T represents the total kinetic energy of the system

$$T = \frac{1}{2}\, m_1\mathbf{v}_1^{\,2} + \frac{1}{2}\, m_2\mathbf{v}_2^{\,2} \tag{11-16}$$

and the potential energy V is that associated with $\mathbf{F}_{12}$. For conservative systems of more than two particles (11-15) continues to be valid provided that T and V are defined in an appropriate way.

An important application of the ideas of energy conservation deals with the relative motion of two *bodies in contact.* Here we must distinguish two cases. First, if the surfaces are rough so that the forces the two bodies exert on each other have a component along their direction of motion, then in general the energy of *neither* body is conserved. Consider, for example, the situation in Figure 11-15, where block A slides at the instantaneous velocity $\mathbf{v}_1$ along a plank B, which in turn is free to move along a *smooth,* horizontal surface. Because the contacts are rough, the force that B exerts on A has a component

$\mathbf{f}$ directed opposite to $\mathbf{v}_1$, and thus the kinetic energy of A decreases. In a similar way, assuming that B is originally traveling to the left, at a speed v_2, so that say the center of mass of A and B is stationary, then it is evident that even though the plank is free to slide on a smooth surface, its kinetic energy also decreases.

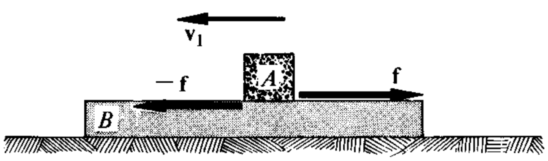


Figure 11-15

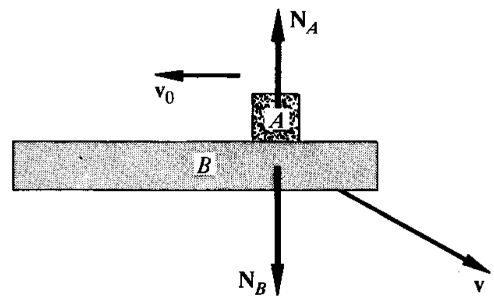


Figure 11-16

The second case of interest occurs when the two surfaces are smooth. Here the total energy of the two bodies is a constant of the motion. To see this, consider in Figure 11-16 a body B undergoing motion under the action of certain conservative forces (not shown) and having the instantaneous velocity $\mathbf{v}$. Suppose that a second body A is also acted upon by conservative forces, so that it slides along B at a velocity $\mathbf{v}_0$. In addition to these external, conservative forces, A and B will also exert forces on each other; let us call them $\mathbf{N}_A$ and $\mathbf{N}_B$, respectively. Because the contacts are presumed smooth, $\mathbf{N}_A$ and $\mathbf{N}_B$ are perpendicular to the surfaces, and thus neither carries out work on the other while they are in relative motion. However, and this is a very important point, since the total system has also the instantaneous velocity $\mathbf{v}$ and since $\mathbf{N}_A$ and $\mathbf{N}_B$ are not generally perpendicular to $\mathbf{v}$, it follows that *they do carry out work on A and B, respectively* as a result of this motion. In other words, the individual energies of the two bodies are *not* conserved. However, since $\mathbf{N}_A = -\mathbf{N}_B$ by virtue of Newton's third law, it follows that the *sum* $(\mathbf{N}_A + \mathbf{N}_B)$ of these two normal forces vanishes and thus does no work on the bodies as they move. Therefore, even though the individual energies of the two bodies are not conserved, the sum of their energies is nevertheless a constant of the motion.

Example 11-7 A block of mass m is released from rest at the top of a smooth, inclined plane of length l, mass M, and angle α (see Figure 11-17a). Suppose that the inclined plane lies on a smooth surface, and is itself free to move in a horizontal direction. Calculate the velocity of the block relative to the plane at the instant when it has traveled a distance s down the plane.

Solution Figure 11-17b shows the situation at a subsequent time t. Let s represent the distance down the plane the block has moved and let $\mathbf{v}_s$ be its velocity at this instant along the plane. As the block slides down under the action of gravity, the inclined plane itself will move to the right; let $\mathbf{v}$ represent this velocity of the inclined plane. The actual velocity of the block consists of the vector sum of its velocity $\mathbf{v}_s$ directed down the inclined plane and the velocity $\mathbf{v}$ of the inclined plane itself.

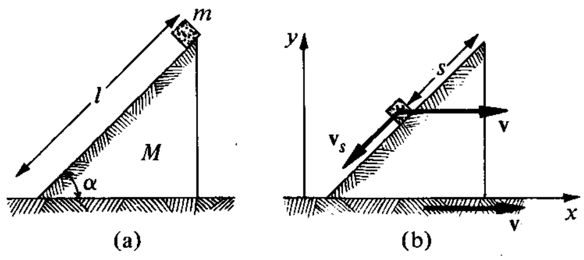


Figure 11-17

Since the external forces have a zero component along the horizontal, x-direction, V_{cx} is constant and must have its initial value of zero for all times t. Substitution into the x-component of (11-14) yields

$$0 = V_{cx} = \frac{Mv + m(v - v_s \cos \alpha)}{M + m}$$

since the velocity of the block along the x-direction is $(v - v_s \cos \alpha)$. Solving for v, we obtain

$$v = v_s \frac{m \cos \alpha}{M + m}$$

A second relation involving the velocities v and v_s can be obtained by use of the fact that the total energy of the plane and the block moving in the gravitational field of the earth is conserved. Even though the normal force that the inclined plane exerts on the block is perpendicular to its surface, the plane is itself in motion and therefore the velocity of the block is not at right angles to this normal force. Thus the energy of the block alone is not conserved. Similarly, the equal and opposite force that the block exerts on the plane carries out work, and thus the energy of the inclined plane alone is not conserved either. However, the total work that these two normal forces carry out vanishes. Hence, since the only external force here is that of gravity, and since this is conservative, the law of conservation of energy enables us to write

$$mgl \sin \alpha = \frac{1}{2} Mv^2 + \frac{1}{2} m[(v - v_s \cos \alpha)^2 + (v_s \sin \alpha)^2] + mgl(l - s) \sin \alpha$$

provided the zero potential energy is taken at the bottom of the plane.

Eliminating the velocity v between these two relations, we obtain

$$v_s^2 = 2g \sin \alpha \left[\frac{M + m}{M + m \sin^2 \alpha}\right] s$$

11-6 Torque and angular momentum—Theorem II

The purpose of this section is to show that (10-9) which relates the torque $\boldsymbol{\tau}$ acting on a particle to the rate of change of its angular momentum (both measured with respect to the same Newtonian origin) is also valid for a system consisting of more than one particle. The precise form of this generalization is encompassed in the following theorem:

Theorem II

> Suppose that the internal forces acting between any pair of particles of a system lie along the line joining them. Then the sum of the torques $\boldsymbol{\tau}$ produced by the *external forces* acting on the system is equal to the time rate of change of $\mathbf{L}$, the sum of the angular momenta of the particles, provided that the common origin about which the torques and the angular momenta are taken is either (1) Newtonian or (2) the center of mass of the system.

In mathematical terms, this theorem is usually expressed by the formula

$$\boldsymbol{\tau} = \frac{d\mathbf{L}}{dt} \tag{11-17}$$

In other words, provided that the symbol $\boldsymbol{\tau}$ represents the sum of the torques produced by all the external forces and that $\mathbf{L}$ represents the sum of the angular momenta of all the particles, then (10-9) is also applicable for systems of more than one particle. Note the hypothesis that the force between any two particles must lie along the line joining them. It no longer suffices to assume, as for Theorem I, that the forces only satisfy Newton's third law. However, for the physical systems of interest to us, the interparticle forces involved generally do lie along the line joining the particles, and therefore this assumption in no way restricts the applicability of the theorem.

It should be emphasized that in (11-17) the symbol $\boldsymbol{\tau}$ represents the torque produced on the particles by the *external* forces only. Just as for Theorem I, we need not concern ourselves with interparticle forces; it is as if there were no internal forces acting at all. Note also that (11-17) is valid for two types of origins: first, if $\boldsymbol{\tau}$ and $\mathbf{L}$ are defined with respect to a Newtonian origin; and second, if they are defined with respect to an origin located at the center of mass of the system.

To establish the validity of Theorem II, let us simplify matters by specializing to the case of a two-particle system. Let us assume further that the origin with respect to which torque and angular momentum are taken is Newtonian so that Newton's laws of motion are valid. The more general derivation necessary to treat the case of more than two particles will be left as an exercise. A proof of the validity of the theorem for the case where the origin is at the center of mass is presented in Example 11-8.

Figure 11-18 shows two particles of masses m_1 and m_2, with instantaneous positions and velocities relative to a Newtonian origin O given by $\mathbf{r}_1$, $\mathbf{v}_1$, and $\mathbf{r}_2$, $\mathbf{v}_2$, respectively. Let $\mathbf{F}_1$ and $\mathbf{F}_2$ represent the external forces acting on the particles and $\mathbf{F}_{12}$ and $\mathbf{F}_{21}$ the internal forces, which we assume act along the line joining them. Making use of the definition of angular momentum in (10-7) and defining the total angular momentum of the two particles in Figure 11-18 by

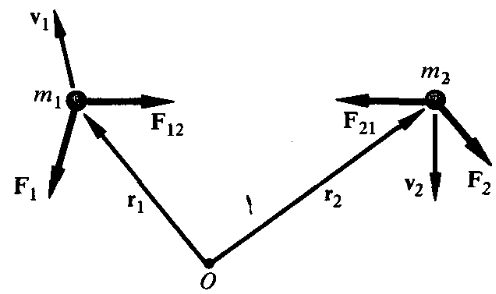


Figure 11-18

the symbol $\mathbf{L}$, we have

$$\begin{aligned}\frac{d\mathbf{L}}{dt} &= \frac{d}{dt}(\mathbf{L}_1 + \mathbf{L}_2)\\ &= \frac{d}{dt}(m_1\mathbf{r}_1 \times \mathbf{v}_1 + m_2\mathbf{r}_2 \times \mathbf{v}_2) \qquad \textbf{(11-18)}\\ &= m_1\mathbf{r}_1 \times \frac{d\mathbf{v}_1}{dt} + m_2\mathbf{r}_2 \times \frac{d\mathbf{v}_2}{dt}\end{aligned}$$

where the third equality follows by use of the fact that the cross product of parallel vectors vanishes. By hypothesis, the origin O is Newtonian, and therefore the equations of motion in (11-6) are applicable. Substituting for the factors $m_1\, d\mathbf{v}_1/dt$ and $m_2\, d\mathbf{v}_2/dt$ by use of these relations we obtain

$$\begin{aligned}\frac{d\mathbf{L}}{dt} &= \mathbf{r}_1 \times (\mathbf{F}_1 + \mathbf{F}_{12}) + \mathbf{r}_2 \times (\mathbf{F}_2 + \mathbf{F}_{21})\\ &= \boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 + (\mathbf{r}_1 - \mathbf{r}_2) \times \mathbf{F}_{12} \qquad \textbf{(11-19)}\end{aligned}$$

where $\boldsymbol{\tau}_1 = \mathbf{r}_1 \times \mathbf{F}_1$ is the torque produced by the external force on particle 1, and similarly for $\boldsymbol{\tau}_2$. Also, in writing the second equality we have used the fact that, according to Newton's third law, $\mathbf{F}_{21} = -\mathbf{F}_{12}$. By hypothesis, the forces between the particles lie along the line joining them, and therefore the cross product $(\mathbf{r}_1 - \mathbf{r}_2) \times \mathbf{F}_{12}$ of the two parallel vectors $(\mathbf{r}_1 - \mathbf{r}_2)$ and $\mathbf{F}_{12}$ vanishes. Finally, since $\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 = \boldsymbol{\tau}$ is the total torque due to external forces acting on the system, the validity of (11-17) is established for this special case.

Example 11-8 Show that (11-17) is also valid for a two-particle system if the origin lies at the center of mass.

Solution The essential difference between this and the case of the Newtonian origin is that the accelerations of the two particles are not given by (11-6). Because of the possibility that the center of mass of the system may be accelerating, these relations are now (see (4-26))

$$m_1\frac{d\mathbf{v}_1}{dt} = \mathbf{F}_1 + \mathbf{F}_{12} - m_1\frac{d\mathbf{V}_c}{dt}$$

$$m_2\frac{d\mathbf{v}_2}{dt} = \mathbf{F}_2 + \mathbf{F}_{21} - m_2\frac{d\mathbf{V}_c}{dt}$$

where $\mathbf{V}_c$ is the velocity of the center of mass. If $\mathbf{r}_1$ and $\mathbf{r}_2$ represent the positions of the particles with respect to the center of mass, so that $m_1\mathbf{r}_1 + m_2\mathbf{r}_2 = 0$, then the steps outlined in (11-18) are still applicable. However, the quantity after the last equality in (11-19) for this case assumes form

$$\begin{aligned}\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 + (\mathbf{r}_1 - \mathbf{r}_2) \times \mathbf{F}_{12} - m_1\mathbf{r}_1 \times \frac{d\mathbf{V}_c}{dt} - m_2\mathbf{r}_2 \times \frac{d\mathbf{V}_c}{dt} \\ = \boldsymbol{\tau}_1 + \boldsymbol{\tau}_2 - (m_1\mathbf{r}_1 + m_2\mathbf{r}_2) \times \frac{d\mathbf{V}_c}{dt}\end{aligned}$$

where again the term $(\mathbf{r}_1 - \mathbf{r}_2) \times \mathbf{F}_{12}$ vanishes for the same reason as before. The validity of (11-17) now follows, since the additional term here

$$-(m_1\mathbf{r}_1 + m_2\mathbf{r}_2) \times \frac{d\mathbf{V}_c}{dt}$$

also vanishes, because the origin has been chosen to lie at the center of mass.[1]

Example 11-9 Two particles of respective masses m_1 and m_2 are connected to the ends of a massless rod and move in the uniform gravitational field $\mathbf{g}$ of the earth.

(a) What is the acceleration of their center of mass relative to a Newtonian origin O?

(b) Show that the angular momentum of the two particles about their center of mass is a constant of the motion.

Solution The situation is shown in Figure 11-19. Let O be a Newtonian origin and let $\mathbf{R}_c$ be the location of the center of mass relative to this origin. Also, let us define the two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ to be the locations of the two particles *relative to the mass center of the system*, so that $m_1\mathbf{r}_1 + m_2\mathbf{r}_2 = 0$.

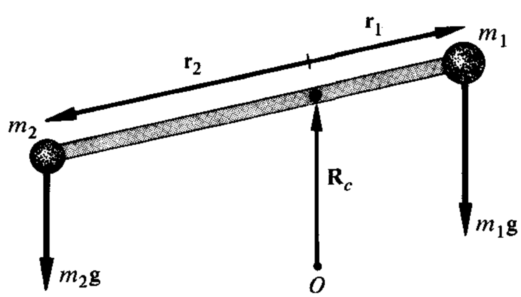


Figure 11-19

(a) Substitution into (11-13), yields

$$(m_1 + m_2)\frac{d\mathbf{V}_c}{dt} = m_1\mathbf{g} + m_2\mathbf{g}$$

where $\mathbf{V}_c$ is the velocity of the mass center relative to O. Canceling out the common

[1]This proof also makes evident the fact that Theorem II is applicable if the acceleration $d\mathbf{V}_c/dt$ of the mass center goes through the origin. In this case, $d\mathbf{V}_c/dt$ and $(m_1\mathbf{r}_1 + m_2\mathbf{r}_2)$ are parallel vectors.

factor $(m_1 + m_2)$, we obtain

$$\frac{d\mathbf{V}_c}{dt} = \mathbf{g}$$

and thus the center of mass has the acceleration of gravity $\mathbf{g}$.

(b) Since the only external forces acting on the two particles are the forces $m_1\mathbf{g}$ and $m_2\mathbf{g}$, the total external torque $\boldsymbol{\tau}$ with respect to the center of mass is

$$\begin{aligned}\boldsymbol{\tau} &= \mathbf{r}_1 \times m_1\mathbf{g} + \mathbf{r}_2 \times m_2\mathbf{g} \\ &= (m_1\mathbf{r}_1 + m_2\mathbf{r}_2) \times \mathbf{g} \\ &= 0\end{aligned}$$

where the third equality follows since $(m_1\mathbf{r}_1 + m_2\mathbf{r}_2)$ vanishes. Hence the total external torque about the center of mass vanishes and therefore, according to Theorem II, the total angular momentum about this origin is a constant.

Generalizing this argument, we find that the angular momentum about the mass center of any system of particles in a uniform gravitational field is a constant of the motion.

11-7 Statics of a rigid body

A *rigid body* is defined as one whose constituents undergo no motion relative to each other under normal circumstances. Strictly speaking, because of the thermal motions of the constituent atoms of a solid, rigid bodies do not truly exist in nature and the concept itself is very much an idealized one. Nevertheless, for some purposes, objects such as a stone, a block of ice, or a meterstick approximate rigid bodies to a high degree.

If the external forces acting on a rigid body are such that it undergoes no rotations or translations of any type, then the body is said to be in a state of *static equilibrium* or in *equilibrium*. In this section we shall study this problem of a rigid body in equilibrium by use of Theorems I and II.

Consider, in Figure 11-20, a rigid body in static equilibrium under the action of certain forces so that it undergoes no translations or rotations of any type. To be specific, suppose that the body is maintained in this state by three external forces $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$, which act at the points of the body described by

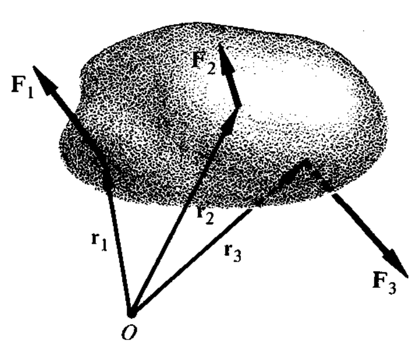


Figure 11-20

the respective position vectors $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ relative to a Newtonian origin O. The fact that only three forces are assumed to act is not essential; the following arguments are easily generalized to the case of more than three.

Let us apply Theorems I and II to this body. Since it is at rest, its center of mass has zero acceleration, and Theorem I implies then that

$$\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = 0 \tag{11-20}$$

Hence in order for a rigid body to be in static equilibrium it is necessary that the vector sum of the external forces vanish. Note that (11-20) is formally identical to (5-2) for a particle. The basic distinction is that the various forces acting on a rigid body will not, in general, act at a single point of the body.

Let us now apply Theorem II to the same body. Again, since the body undergoes no motion it follows that the sum of the angular momenta of its constituents vanishes for all times. Hence, by Theorem II,

$$\mathbf{r}_1 \times \mathbf{F}_1 + \mathbf{r}_2 \times \mathbf{F}_2 + \mathbf{r}_3 \times \mathbf{F}_3 = 0 \tag{11-21}$$

which states that the sum of the external torques about the fixed origin must vanish. The relations in (11-20) and (11-21) are the necessary conditions that must be satisfied in order for a rigid body to be in static equilibrium.

From a practical point of view, (11-20) and (11-21) represent six linear relations among the nine components of the external forces $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$. That is, if these three vectors are resolved into components along an arbitrary set of coordinate axes, then (11-20) and (11-21) represent six linear relations among these nine components. If any three of these nine are known, then the remaining six may be determined.

More generally, if a rigid body is in a state of static equilibrium under the action of N forces $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_N$, then the appropriate generalization of (11-20) and (11-21) are six relations among the $3N$ components of these forces. Thus if $(3N - 6)$ of these are known, the remaining six can be determined. The impossibility of obtaining additional, linearly independent relations by calculating torques about a different origin is established in the problems.

Example 11-10 A uniform ladder of mass m and of length l leans against a smooth, vertical wall and is in contact with a rough floor at an angle of 45°, as shown in Figure 11-21. Calculate the forces that the wall and the floor exert on the ladder, assuming that it is in static equilibrium.

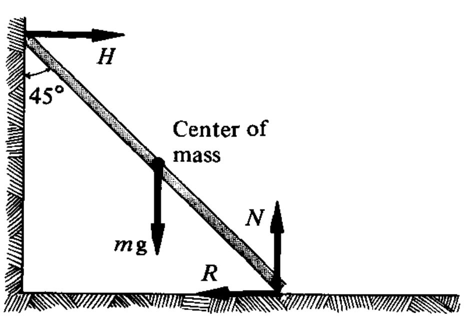


Figure 11-21

Solution In addition to the force of gravity $m\mathbf{g}$, which may be taken to act vertically downward at the center of mass, there is a horizontal force, call it **H**, on the ladder at the upper end, and at the lower end there is a force which is conveniently resolved into a vertical component N and a horizontal component R.

First, since there is no motion of the center of mass along the vertical or the horizontal direction, it follows from (11-20) that

$$N - mg = 0 \qquad H - R = 0$$

A third relation, involving the three unknowns N, H, and R, can be obtained by taking torques about the bottom of the ladder. The torque that gravity produces about this point has the magnitude $mg(l/2)\cos 45°$ since the perpendicular distance from the origin to the line of action of this force is $(l/2)\cos 45°$. Similarly, H produces about the same origin the torque $Hl\cos 45°$. The torque produced by the gravitational force is perpendicular to and out of the plane of the page in Figure 11-21, whereas the torque produced by **H** is directed perpendicularly down into this plane. The force produced by the floor on the ladder produces no torque about this origin. Substitution into (11-21) yields

$$Hl\cos 45° = mg\left(\frac{l}{2}\right)\cos 45°$$

and solving for the unknowns we find

$$H = R = \frac{1}{2}mg \qquad N = mg$$

Example 11-11 Repeat Example 11-10, but assume this time that the wall is not smooth so that in addition there is now a vertical component V to the force produced on the ladder by the wall.

Solution This time there are four unknowns to be determined, namely N, R, H, and V. Two equations relating these are obtained by the requirement that the components of the forces acting along the horizontal and vertical directions add to zero:

$$N + V - mg = 0$$
$$H - R = 0$$

A further relation involving these may be obtained by taking torques about some fixed point, say the top or the bottom of the ladder. Thus we obtain a third relation among these four unknowns. Since there are no further relations independent of these three, it follows that the problem stated is indeterminate.

Example 11-12 A sphere of mass m and radius a is held in a state of static equilibrium on an inclined plane of angle α by a fixed vertical string attached to the sphere (see Figure 11-22). Calculate the tension T in the string and the force that the plane exerts on the sphere.

Solution Let R and N be the components along and perpendicular to the plane of the force which the plane exerts on the sphere. In addition to these, the only other forces acting are the downward pull of gravity mg at the center of the sphere and the vertically upward tension T in the string. Since the center of mass of the sphere undergoes no motion, it follows that the sum of the forces along and perpendicular to

the inclined plane vanish. Thus

$$
\begin{aligned}
R + T \sin \alpha - mg \sin \alpha &= 0 \\
N + T \cos \alpha - mg \cos \alpha &= 0
\end{aligned} \tag{11-22}
$$

To obtain a third relation involving R, T, and N, let us take torques about that point on the sphere where the string is attached. N and T produce no torques about this point, and the remaining two forces R and $m\mathbf{g}$ produce the net torque $(2aR - mga \sin \alpha)$, directed perpendicular to and out of the plane of Figure 11-22. The requirement that this torque vanish leads to

$$R = \frac{1}{2} mg \sin \alpha$$

and substitution into (11-22) yields

$$T = \frac{1}{2} mg$$

$$N = \frac{1}{2} mg \cos \alpha$$

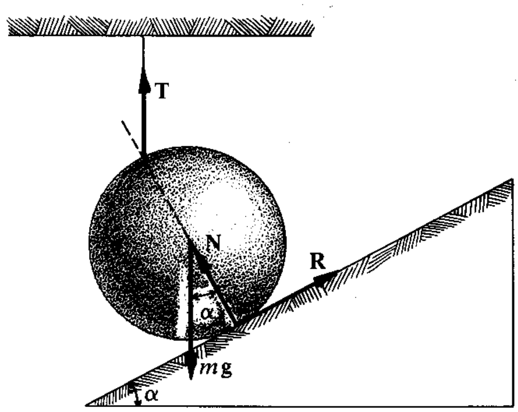


Figure 11-22

†11-8 Relative motion of a two-body system

In this section we shall establish that the laws that govern the *relative* motion of the members of an isolated two-body system is the same as that of a single particle moving in an external force field.

Consider, in Figure 11-23, two isolated particles of masses m_1 and m_2 and let their instantaneous positions and velocities with respect to a fixed Newtonian origin 0 be $\mathbf{r}_1$, $\mathbf{v}_1$ and $\mathbf{r}_2$, $\mathbf{v}_2$, respectively. Since the particles are isolated, the equations of motion in (11-6) become

$$
\begin{aligned}
m_1 \frac{d\mathbf{v}_1}{dt} &= \mathbf{F}_{12} \\
m_2 \frac{d\mathbf{v}_2}{dt} &= \mathbf{F}_{21}
\end{aligned} \tag{11-23}
$$

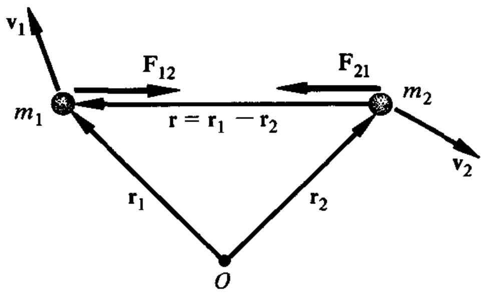


Figure 11-23

Dividing the first by m_1 and the second by m_2, and subtracting them we obtain

$$\frac{d}{dt}(\mathbf{v}_1-\mathbf{v}_2)=\mathbf{F}_{12}\left(\frac{1}{m_1}+\frac{1}{m_2}\right) \tag{11-24}$$

since, according to (11-2), $\mathbf{F}_{21}=-\mathbf{F}_{12}$.

In order to see the significance of this relation, let us define the *reduced mass* μ of this system by

$$\frac{1}{\mu}=\frac{1}{m_1}+\frac{1}{m_2} \tag{11-25}$$

and the *relative coordinate* $\mathbf{r}$ by

$$\mathbf{r}=\mathbf{r}_1-\mathbf{r}_2 \tag{11-26}$$

As shown in the figure, $\mathbf{r}$ is the position of m_1 as seen by an observer fixed with respect to m_2. The time derivative of $\mathbf{r}$ is the relative velocity

$$\mathbf{v}=\frac{d}{dt}(\mathbf{r}_1-\mathbf{r}_2)=\mathbf{v}_1-\mathbf{v}_2 \tag{11-27}$$

which represents the velocity of m_1 as seen by an observer at rest relative to m_2. Substituting (11-25) and (11-27) into (11-24), we obtain

$$\mu\frac{d\mathbf{v}}{dt}=\mathbf{F}_{12} \tag{11-28}$$

which may be stated in words as follows:

The motion of m_1, as seen by an observer fixed with respect to m_2, is as if m_1 were replaced by a particle of reduced mass μ in (11-25) *and moving in the given force field* $\mathbf{F}_{12}$.

For example, to describe the dynamics of the relative motion of the earth and the sun, we may take an origin fixed at the center of the sun and then describe the motion of the earth relative to this origin, provided that the mass of the earth is replaced by the reduced mass of the earth and the sun. Hence the calculations of the motion of the earth about the sun in Chapter 10 are fully correct if the mass m of the earth in these formulas is replaced by the reduced

mass $\mu = mM/(m+M)$, where M is a solar mass. However, since $m/M \approx 10^{-6}$, it follows that, correct to one part in a million, the reduced mass of the earth-sun system is the same as that of the earth. In other words, the earth and the sun do indeed orbit about their common center of mass, but since $m/M \ll 1$, this origin can be assumed to coincide with the center of the sun to a high degree of accuracy (see Example 9-13).

11-9 Summary of important formulas

There are two basic theorems that characterize the dynamics of a composite system. These are as follows:

Theorem I

$$\mathbf{F} = M\frac{d\mathbf{V}_c}{dt} \tag{11-13}$$

where

$$\mathbf{F} = \sum_i \mathbf{F}_i$$

is the sum of the *external* forces acting on the system of total mass M, and $\mathbf{V}_c$ is the velocity of the center of mass.

Theorem II

$$\boldsymbol{\tau} = \frac{d}{dt}\mathbf{L} \tag{11-17}$$

where $\boldsymbol{\tau}$ is the torque produced by the *external* forces acting on the system and $\mathbf{L}$ is its total angular momentum, provided that the origin with respect to which $\boldsymbol{\tau}$ and $\mathbf{L}$ are defined is either Newtonian or coincides with the center of mass of the system.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** internal forces; **(b)** mass center; **(c)** rigid body; and **(d)** static equilibrium.
2. State Theorem I. What are the essential assumptions made in deriving this theorem?
3. State Theorem II. In what way are the assumptions made in deriving this theorem different from those used in deriving Theorem I?
4. Can the dynamics of a three-body system be described in its entirety by use only of Theorems I and II? Explain.
5. Describe the circumstances (if any) for which the application of Theorems I and II to a two-particle system will describe the dynamics of this system in its entirety.
6. A small boy holds a water hose in such a way that a stream of water

shoots out from the nozzle in a horizontal direction. What path in space do the water droplets follow? Justify your answer.

7. What is the basis of the statement that the center of mass of the entire universe moves with uniform velocity in a straight line?
8. Is the total angular momentum of the universe with respect to its center of mass a constant of the motion? Are there any other origins for which there is a conservation law for the angular momentum of the universe?
9. The cheerleader at a football game throws a baton high into the air. Why is the angular momentum of the baton, about the center of mass, constant during its flight? Why is its angular momentum not constant about any other origin?
10. Suppose that the internal force between two particles did not lie along the line joining them. What peculiar physical effects would you expect to see under this circumstance? Illustrate your answer by reference to a particular case.
11. What is the common characteristic, from the viewpoint of Theorem I, of the paths followed by all divers, regardless of their skills, in diving from a given diving board?
12. What difficulties arise if we attempt to apply Newton's law of motion for a particle to describe the dynamics of a macroscopic body? Under what circumstances are there no difficulties in doing this?
13. Explain in physical terms why the energy associated with the motion of the center of mass of a system is generally not its total energy. Give two examples of energy associated with a compound system that differ from its center-of-mass energy.
14. For an isolated two-body system, we know that the kinetic energy associated with the motion of the center of mass is a constant of the motion. Is this a new constant of the motion or is it related to the conservation-of-momentum law? Explain.
15. Consider the situation in Figure 11-15, and suppose this time that all contacts are smooth and further that B lies on a smooth, horizontal surface. Explain why the kinetic energies of A and of B are separately conserved.
16. Consider the situation in Figure 11-16. Assuming that the contacts between A and B are smooth, explain why if $\mathbf{v}$ is zero and A slides along B with velocity $|\mathbf{v}_0| \neq 0$, the energies of A and B are separately conserved. Explain also why if $\mathbf{v}$ does not vanish, then in general neither the energies of A nor of B are conserved but their sum is.
17. By reference to a particular case, show that even though the sum of the external forces acting on a rigid body vanishes, the body need not be in equilibrium.
18. Suppose that the torques produced by the external forces acting on a rigid body vanish about two distinct origins. Is it necessary that the body be in equilibrium under these circumstances? Explain.
19. If the word "particles" is replaced by the word "bodies" in the statement of Theorem I, under what conditions will this theorem be valid? Justify your answer.
20. If the two particles in Figure 11-19 are falling in an inverse-square gravitational field, instead of a uniform field, is the torque about the center of mass of the particles still a constant of the motion? For what two positions of the massless rod will there be no torque about this mass center?
21. Two particles of masses m_1 and m_2 are observed to travel along the respective trajectories $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$. Devise an operational procedure, in

terms of these trajectories, to determine whether or not these particles exert forces on each other. Does the fact that the motions of the particles may be correlated suffice to enable us to conclude that the interparticle forces are not zero? Explain.

PROBLEMS

1. Three particles of masses 1 kg, 2 kg, and 3 kg are located at the vertices of an equilateral triangle having sides 50 cm in length. Find the mass center of this collection, and give its coordinates in terms of a system with its origin at the 1 kg particle and with the 2 kg particle located along the positive x-axis.
2. Small weights having masses of 50 grams, 100 grams, and 75 grams are attached to places along a meter stick corresponding to the respective readings: 0 cm, 70 cm, and 100 cm. What is the location of the mass center of this arrangement, assuming that the mass of the meterstick is negligible?
3. Repeat Problem 2, but assume this time that the meterstick is uniform and has a mass of 100 grams.
4. Three rods having respective masses of 300, 400, and 500 grams are uniform and thus have lengths proportional to their masses. If the shortest has a length of 30 cm and if they are joined together to form a triangle, find the perpendicular distance from the center of mass to each of the two shorter rods.
5. A rod of length l has a mass per unit length, λ, that increases linearly with distance from one end. If its total mass is M and its mass per unit length at the lighter end is λ_0, find the distance of the center of mass from the lighter end.
6. (a) Show that the center of mass of a three-particle system is the same as that of the "two-particle system" consisting of (1) one of the particles itself, and (2) a second particle located at the mass center of the two remaining ones and having a mass equal to their sum.
   (b) Show by use of your result to (a) that the mass center of a system consisting of a single particle, call it A, and an arbitrary collection of particles, call it C, is the same as that of the two-particle system consisting of A and a particle with mass the same as that of C and located at the center of mass of C.
7. Making use of the results of Problem 6, prove that if $\mathbf{R}_A$ and $\mathbf{R}_B$ are the mass centers of two *bodies* of masses M_A and M_B, respectively, then the center of mass $\mathbf{R}_c$ of the total system is

$$\mathbf{R}_c = \frac{M_A \mathbf{R}_A + M_B \mathbf{R}_B}{M_A + M_B}$$

8. A circular disk of mass $M = 500$ grams is pulled along a smooth, horizontal surface by a force $F_0 = 2N$ applied at a point on its rim.
   (a) If the line of action of $\mathbf{F}_0$ goes through the center of the disk, as in Figure 11-24a, what is the acceleration of the center of mass of the disk?

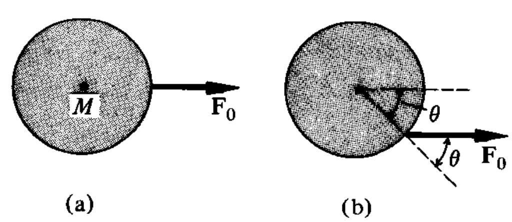


Figure 11-24

(b) For the situation shown in Figure 11-24b, where the direction of $\mathbf{F}_0$ makes an angle θ with the radius, what is now the acceleration of the center of mass?
(c) Contrast the motion of the disk in these two cases.

9. Repeat (a) and (b) of Problem 8, but assume this time that the disk lies on a rough, horizontal surface, which is characterized by a coefficient of sliding friction $\mu = 0.1$.
10. Consider the physical system in Figure 11-5 and suppose that $m_1 = 3$ kg and $m_2 = 1$ kg, and that just prior to the initial instant when the constant force $F_0 = 0.5$ newton is applied both bodies are at rest. **(a)** Calculate the acceleration of the mass center. **(b)** Find the displacement of the mass center at time t.
11. Consider the situation in Problem 10, but suppose this time that at the instant when $\mathbf{F}_0$ is applied, m_2 is at rest and m_1 has the velocity of 2 m/s, directed to the left. Calculate now the displacement of the mass center at time t. Does it matter whether the spring was stretched or unstretched initially?
12. Consider the physical situation described in Figure 11-14. Suppose that all the particles are identical and have the same mass of 1.5 kg and that $F_0 = 3$ newtons. Assuming that all four particles are at rest at the instant just prior to that when $\mathbf{F}_0$ is applied, calculate the displacement of the mass center from its initial position at time t. What would be the displacement of the *mass center* at time t if at a time τ $(< t)$ the three springs connected to the topmost particle snapped?
13. A 1 kg particle moves along the trajectory $\mathbf{r}_1(t) = 2t\mathbf{i} + t^3\mathbf{j}$, where t is in seconds and all lengths are in meters. In addition, suppose that a 2 kg particle travels along a trajectory $\mathbf{r}_2(t) = (3 + 4t^2)\mathbf{i} - t^3\mathbf{j}$, where, as above, $\mathbf{i}$ and $\mathbf{j}$ are unit vectors in a certain coordinate system.
    (a) Calculate the force on each particle at time t.
    (b) Calculate the acceleration of the center of mass.
    (c) By use of your result to (b), calculate the total external force acting on the system at time t. Should this be the same as the vector sum of the forces calculated in (a)?
14. In a certain coordinate system, a particle of mass 1 kg moves along the trajectory $\mathbf{r}_1(t) = \mathbf{i}\cos t + \mathbf{j}\sin t$, where t is in seconds and all lengths are measured in meters. Suppose in addition that there is a second, identical particle, which moves in a way so that the mass center of the combined system has the trajectory

$$\mathbf{R}_c = 2\mathbf{i}\cos t + 2\mathbf{j}\sin t$$

    (a) Determine the trajectory of the other particle.
    (b) Calculate the total force acting on the particles.
15. Two identical particles, each of mass m, move in the interior of a smooth, horizontal, circular tube having a radius of R_0, as shown in Figure 11-25. Suppose that initially the particles are located at the points $(R_0, 0)$ and $(-R_0, 0)$ in the coordinate system shown in the figure and have the velocity $\mathbf{v}_0$. For times t prior to

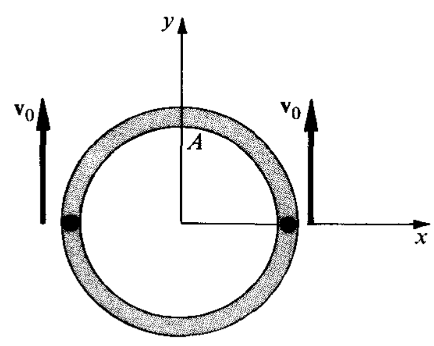


Figure 11-25

that when the particles collide at the point A:

(a) Show that the center-of-mass coordinate is

$$\mathbf{R}_c = \mathbf{j}R_0 \sin \frac{v_0 t}{R_0}$$

(b) Calculate the sum of the external forces acting on the two particles.

(c) Calculate the force acting on each particle separately.

(d) Compute the sum of the forces calculated in (c) and compare them with your result to (b).

16. A projectile is fired from a cannon with an initial velocity $\mathbf{v}_0$ at an angle α with respect to horizontal. At the height of its trajectory the projectile explodes into two equal parts. If an observer on the ground notes that one of these fragments goes vertically upward immediately after the explosion at a speed v_1, what is the subsequent trajectory of (a) the center of mass and (b) the other fragment. Express your answer in terms of the coordinate system in Figure 11-7.

17. Consider, in Figure 11-26, two blocks of respective masses m_1 and m_2, connected by a spring of natural length l and of constant k and confined to one-dimensional motion along a smooth, horizontal line.

(a) Show that the equation of motion for m_1 is

$$m_1 \frac{dv_1}{dt} = -k[l - (x_2 - x_1)]$$

where $v_1 = dx_1/dt$ is the velocity of m_1.

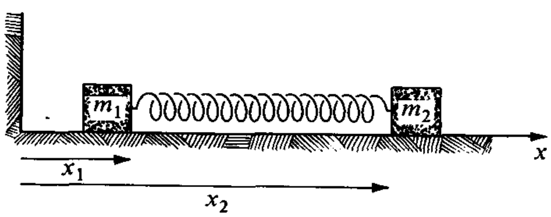


Figure 11-26

(b) Find the corresponding equation of motion for m_2.

(c) Check the consistency of the above results by calculating the acceleration of the center of mass of the two bodies.

*18. Suppose that the force between two isolated particles is derivable from a potential V_{12} which is a function only of the distance $|\mathbf{r}_1 - \mathbf{r}_2|$ separating the particles. Show that the forces $\mathbf{F}_{12}$ and $\mathbf{F}_{21}$ derived from this potential satisfy Newton's third law.

*19. Suppose that the sum of the external forces $\mathbf{F}$ acting on a system are conservative. Prove that the quantity

$$\frac{1}{2} M\mathbf{V}_c^2 + V(\mathbf{R}_c)$$

where $V(\mathbf{R}_c)$ is the potential energy function associated with $\mathbf{F}$ is a constant of the motion.

*20. Consider an isolated two-particle system of masses m_1 and m_2 and suppose that the force $\mathbf{F}_{12}$ between the particles is derivable from the potential function $V = V(|\mathbf{r}_1 - \mathbf{r}_2|)$.

(a) Why is the quantity E_c defined by

$$E_c = \frac{1}{2}(m_1 + m_2)\mathbf{V}_c^2$$

where $\mathbf{V}_c$ is the velocity of the mass center of the particles, a constant of the motion?

(b) Show that the quantity E defined by

$$E = \frac{1}{2}(m_1\mathbf{v}_1^2 + m_2\mathbf{v}_2^2) + V(|\mathbf{r}_1 - \mathbf{r}_2|)$$

is constant in time.

(c) Show that the constant of the motion $(E - E_c)$ can be expressed entirely in terms of the relative coordinate $(\mathbf{r}_1 - \mathbf{r}_2)$ and the relative velocity $(\mathbf{v}_1 - \mathbf{v}_2)$.

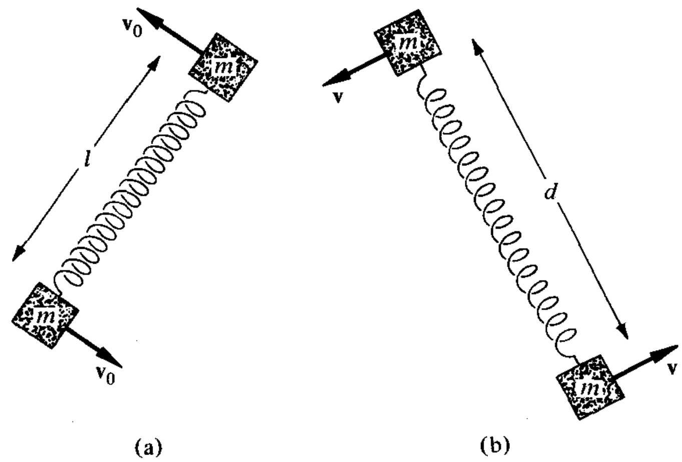


Figure 11-27

21. Figure 11-27 shows (top view) two blocks, each of mass m and connected by a spring of natural length l and constant k and lying in equilibrium on a smooth and horizontal surface. Suppose that suddenly they each receive an impulsive blow, so they start to travel with the velocity $\mathbf{v}_0$ perpendicular to the line joining them, as shown in Figure 11-27a.
    (a) What is the velocity of the center of mass?
    (b) A short time later, the blocks are observed to be a distance d apart and traveling instantaneously at the velocity v perpendicular to the spring (see Figure 11-27b). Show that

$$vd = v_0 l$$

    (c) Obtain a second relation between v, d, v_0, and l by use of the law of energy conservation.

*22. By use of the result in Example 11-7, show that the acceleration of the inclined plane in Figure 11-17 is

$$a = \frac{mg \sin 2\alpha}{2(M + m \sin^2 \alpha)}.$$

and also calculate the acceleration a_s of the particle down the inclined plane.

23. Making use of the result of Problem 22, calculate (a) the normal force N that the plane in Figure 11-17 exerts on the particle and (b) the force that the floor exerts on the inclined plane. Recall that all contacts are presumed to be smooth.

*24. A particle of mass m is attached by means of a massless string to a block of mass M, which in turn is free to slide along a horizontal and smooth rail. Suppose that initially the particle is displaced from the vertical by an angle θ_0 and released from rest. Figure 11-28 shows the situation at a certain subsequent instant when the block has a velocity $\mathbf{v}$ directed to the right and the particle has a certain tangential velocity $\mathbf{v}_\theta$ with respect to the suspension point of the block.
    (a) Show that since the external forces acting on the system have no component along the horizontal direction, then

$$Mv + m(v_\theta \cos \theta + v) = 0$$

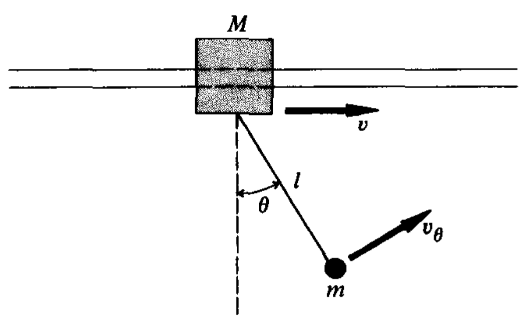


Figure 11-28

(b) Why must the quantity

$$E = \tfrac{1}{2}Mv^2 + \tfrac{1}{2}m[(v + v_\theta \cos\theta)^2 + (v_\theta \sin\theta)^2] - mgl\cos\theta$$

be a constant of the motion? Using the given initial values, find a second relation between v and v_θ, and find the speed v of the block when $\theta = 0$.

25. Suppose that (11-20) and (11-21) are satisfied for a body in static equilibrium with $\mathbf{r}_i$ $(i = 1, 2, 3)$ referred to a given Newtonian origin. Show that (11-21) will automatically be satisfied if torques are taken with respect to any other Newtonian origin. *Hint*: Let $\mathbf{r}_0$ be the displacement between the two origins and let $\mathbf{r}_i' = \mathbf{r}_0 - \mathbf{r}_i$. Show then that if $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_3$ satisfy (11-20) and (11-21), then

$$\sum_{i=1}^{3} \mathbf{r}_i' \times \mathbf{F}_i = 0$$

*26. Show that, as far as translations and rotations about the center of mass are concerned, for a collection of particles in a uniform gravitational field, the external force may be thought of as acting only at the mass center. That is, show that the total force on the system produces no torques on the particles about the center of mass.

27. Suppose that the entire universe consisted only of the earth and the sun and that in addition to the gravitational forces between them there existed a second force on each proportional to its respective mass. In other words, suppose that in addition to the gravitational force between them, each of these bodies experienced a force $\mathbf{F} = m\boldsymbol{\alpha}$, where m is the mass of the earth or the sun and $\boldsymbol{\alpha}$ is a constant vector.
   (a) Show that this external force produces no torques about the center of mass.
   (b) Show further that their relative motion—that is, the motion of the sun as seen by the observer fixed on the earth—is not affected by this external force.
   (c) Can you devise an operational way by means of which you could tell whether or not such an external force actually exists?

*28. Suppose that there existed in the universe a certain force $\mathbf{F}_e$, which acted on all matter and had the form $\mathbf{F}_e = m\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a constant vector and where m is the mass of any body of interest.
   (a) Show that the relative motion of any two particles in the universe would not be affected by such a force. That is, show that the *relative* acceleration between any two particles is not affected by $\mathbf{F}_e$ according to Newton's laws of motion.
   (b) Show that this force produces no torque relative to the mass center of any collection of particles in the universe.
   (c) Can you think of a way that we could determine whether or not such a force actually existed?

29. Two particles of respective masses 3 kg and 2 kg are attached to a massless rod 1.5 meters long. Suppose that there is a hole in the rod at its mass center and that the rod is suspended there by a smooth peg so that it is free to rotate in a vertical plane.
   (a) Calculate the force which the peg exerts on the rod.
   (b) What is the torque, about the mass center, on the rod?
   (c) If the rod is started rotating at an angular velocity of 2 rad/s, what is its angular velocity at any subsequent time t?

30. A uniform rod of length 0.8 meter and of mass 10 kg is kept in a horizontal equilibrium position by two

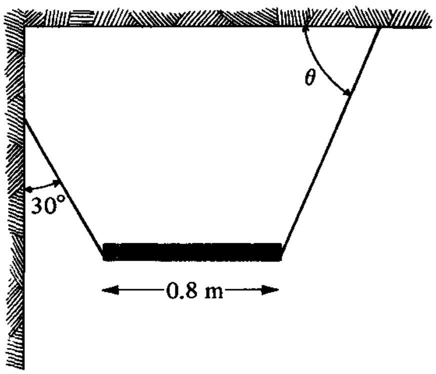


Figure 11-29

strings, as shown in Figure 11-29. Determine the tension of each string and the angle θ that the string on the right makes with the ceiling. Assume that the other string makes an angle of 30° with the vertical wall.

31. A 1.5-meter uniform steel rod has a mass of 20 kg and is attached at one end to a vertical wall at an angle of 60°. As shown in Figure 11-30, the rod is supported at its other end by a horizontal cable and, in turn, supports a block of mass 100 kg from this end. Calculate the tension T in the cable and the horizontal component, (H), and the vertical component (V) of the force that the wall exerts on the rod.

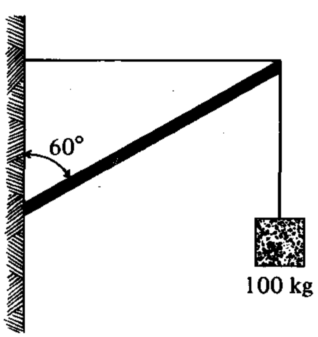


Figure 11-30

32. A uniform sphere of radius a and mass M is kept in equilibrium on an inclined plane of angle α by a horizontal string attached to the highest point of the sphere (see Figure 11-31). Determine, in terms of a, M, and α, the tension in the string and the magnitude and the angle θ (with the horizontal) of the force which the plane exerts on the sphere.

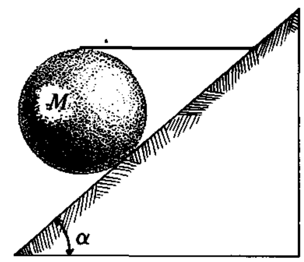


Figure 11-31

33. A sphere of radius 10 cm and mass 10 kg is kept in equilibrium on an inclined plane of angle 30° and by a smooth, vertical wall (see Figure 11-32). Calculate the forces that the two surfaces exert on the sphere.

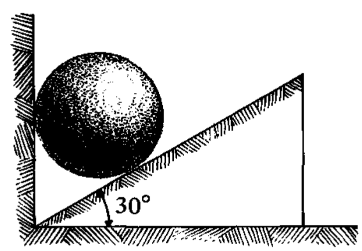


Figure 11-32

34. A uniform 2-meter rod of mass 2 kg is kept in a horizontal equilibrium position by a fulcrum at a distance of 50 cm from one end and by a certain body of mass m hung at that end. Calculate m and the magnitude of the upward force that the fulcrum exerts on the rod.
35. A 5-meter uniform ladder has a mass of 25 kg and leans against a smooth, vertical wall at an angle of 30°. Calculate the horizontal (H) and the vertical (V) components of the force on the ladder produced by the floor and the force F due to the wall.
36. A uniform rod 2.5 meters long has a mass of 15 kg and, as in Figure 11-33,

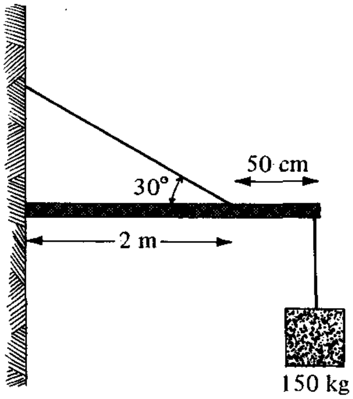


Figure 11-33

is maintained in a horizontal equilibrium position by a light wire attached 50 cm from one end and at an angle of 60° to a vertical wall. Suppose that one end of the rod is attached to the wall and that a 150-kg mass is attached at the other end.

(a) Write down the conditions for equilibrium in terms of the tension T in the wire and the horizontal and the vertical components H and V of the forces that the wall exerts on the rod.

(b) Determine numerical values for T, H, and V.

(c) Repeat (a) and (b), but this time compute the torques about some other point. Do you get the same result? Should you?

†37. Calculate the reduced mass of each of the following systems: (a) earth-moon; (b) electron-proton; and (c) neutron-proton.

†38. Prove the following properties of the reduced mass μ of two particles of masses m_1 and m_2.

(a) If $m_1 = m_2$, then $\mu = \frac{1}{2}m_1$.

(b) If $m_1 \gg m_2$, then $\mu \cong m_2$.

(c) $\mu < m_1$ and $\mu < m_2$ for all m_1 and m_2.

†39. Show that if $\mathbf{r}$ $(\equiv \mathbf{r}_1 - \mathbf{r}_2)$ is the relative position vector of two particles of masses m_1 and m_2 and $\mathbf{R}_c$ is their center of mass, then

$$\mathbf{r}_1 = \mathbf{R}_c + \frac{\mu}{m_1}\mathbf{r}$$

$$\mathbf{r}_2 = \mathbf{R}_c - \frac{\mu}{m_2}\mathbf{r}$$

with μ the reduced mass.

†40. Consider the physical system in Figure 11-26, and suppose that initially the spring is compressed by 5 cm and then both bodies are released at rest.

(a) Find the location of the mass center of any subsequent time t assuming it was initially at the origin.

(b) Write down the equation of motion for the relative coordinate $x = x_2 - x_1$.

(c) Explain how you could determine the positions of the blocks at any time t.

†41. Show that if allowance is made for the fact that the earth and the sun orbit about their center of mass, instead of about the sun as we assumed in Chapter 10, then the semimajor axis a and the period T are related by

$$T = \frac{2\pi a^{3/2}}{[G(M+m)]^{1/2}}$$

†42. Consider the motion of two isolated particles having respective masses of 10 grams and 20 grams. Suppose that, as seen by an observer at rest relative to the 20-gram particle, the other appears to be orbiting about it in a circle of radius 30 cm at a constant velocity of 90 cm/s.

(a) Show that the force on the lighter particle is 0.018 newton.

(b) Show that the orbits of the two particles as seen by an observer with origin at their center of mass are circles. Determine the radii of these circles.

(c) What is the velocity of each particle as seen by the observer in (b)? (*Hint*: Why must the angular velocity ω of the two particles in (a) and (b) be the same?)

12 The motion of a rigid body

Science is built up with facts as a house with stones; but a collection of facts is no more a science than a heap of stones is a house.

JULES HENRI POINCARÉ (1854–1912)

12-1 Introduction

In Chapter 11 we developed two very basic theorems of mechanics. The first of these describes the dynamics of the mass center of a composite system, and the second deals with the rotational dynamics of such a system. In general, these two theorems alone do not make possible a complete description of the dynamical behavior of such systems. However, there is one important exception to this rule; this is the case of a rigid body. Here, as will be discussed in this chapter, Theorems I and II do indeed constitute a complete dynamical description.

It is not hard to see why the translational motion of a rigid body and its rotational motion about the mass center constitute a complete specification of the motion of such a body. For if the position of the mass center and the orientation of the body relative to a fixed set of axes are known, then so is the position of every point of that body. To see this consider, for example, the thin rod in Figure 12-1. Assuming that it lies in the x-y plane, it is apparent that each point of the rod may be expressed directly in terms of the coordinates

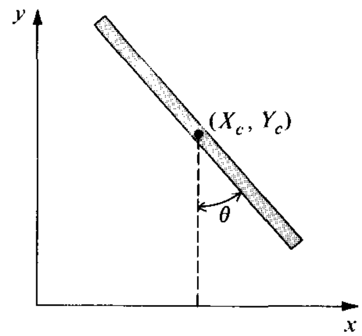


Figure 12-1

(X_c, Y_c) of its center of mass and the angle θ it makes with the y-axis. Similar arguments can be made for any rigid body. Therefore, if the orientation of, and the position of, the center of mass of a rigid body are known at any time t, then, in effect, so is the motion of every point of that body.

An important consequence of this ability to specify the motion of a rigid body in terms only of its orientation and the position of its mass center is that its dynamical behavior is completely determined by the *external* forces and torques acting on it. For if the external forces are known, the position of the mass center may be calculated at any time t by use of Theorem I. Correspondingly, the orientation of the body can be determined in terms of the external torque by applying Theorem II with the center of mass as origin. A rigid body, therefore, is a rather unique physical system in that its dynamical behavior is determined completely by the external forces acting on it.

12-2 The angular momentum of a rigid body

We say that a rigid body *rotates about an axis* at the angular velocity ω provided that every point of the body orbits at this angular velocity in a circle with center at some point on the axis. Thus all the constituents of a rigid body rotating about an axis orbit at the same angular velocity ω but in circles of generally different radii. For some purposes, it is convenient to think of the angular velocity of a rotating body as a vector $\boldsymbol{\omega}$ with magnitude equal to the common angular velocity of the constituents and to be directed parallel to the axis of rotation with sense given by the right-hand rule. That is, if in Figure 12-2 the axis of rotation is grasped in the right hand with the fingers pointing along the direction in which the particles orbit, then the thumb points along $\boldsymbol{\omega}$.

Consider, in Figure 12-3, a rigid body rotating at an angular velocity $\boldsymbol{\omega}$ about a fixed axis. Although it is assumed here that the axis of rotation goes through the body itself, this need not be the case. Suppose m_i is the mass of a typical particle whose instantaneous position vector with respect to an origin O at a point on the axis of rotation is $\mathbf{r}_i$. As the body rotates about this axis, this particle will orbit in a circle whose plane is perpendicular to the axis of rotation and with a

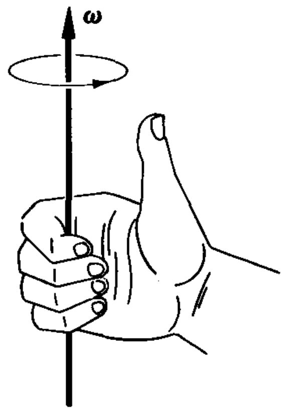


Figure 12-2

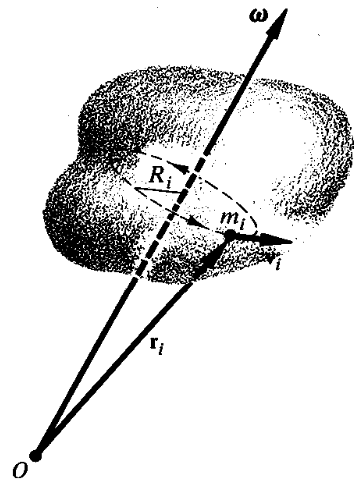


Figure 12-3

radius numerically equal to the component R_i of $\mathbf{r}_i$ perpendicular to the axis. According to (5-15), the velocity $\mathbf{v}_i$ of this particle has the magnitude $v_i = R_i\omega$ and its direction is tangent to the circle on which the particle moves.

The total angular momentum $\mathbf{L}$ of this body is the vector sum of the individual angular momenta of the constituent particles. To simplify matters, let us confine ourselves in the following to those physical situations for which the direction of the axis of rotation does not vary in time. For these, only the component L, of the angular momentum along this axis, is of interest. Accordingly, defining L_i to be the component of the angular momentum $\mathbf{L}_i$ of the ith particle along $\boldsymbol{\omega}$, we find by use of (10-7) that

$$L_i = m_i R_i v_i = m_i R_i^2 \omega \tag{12-1}$$

where the first equality follows since R_i is the component of $\mathbf{r}_i$ perpendicular to $\mathbf{v}_i$ and the second since $v_i = \omega R_i$. Summing (12-1) over all particles, we find

$$L = \sum L_i = \omega \sum m_i R_i^2 \tag{12-2}$$

where, since each constituent particle travels at the same angular velocity, the factor ω could be taken out from under the summation.

Therefore, when a rigid body rotates about a fixed axis, the component L of its angular momentum along that axis is proportional to the angular velocity ω. In mathematical terms, (12-2) may be expressed in the form

$$L = I\omega \tag{12-3}$$

where I is a certain coefficient of proportionality, defined by

$$I = \sum m_i R_i^2 \tag{12-4}$$

This coefficient I plays a very important role in studies of rigid-body motion and is known as the *moment of inertia of the body about the given axis.* According to its definition, the moment of inertia of a rigid body depends on

the geometric structure of that body as well as on the distribution of matter within it. In addition, it also depends on the axis. Recall that in (12-4) R_i represents the perpendicular distance from the ith particle to the axis of rotation and in general R_i will differ for different axes. Physically, the moment of inertia of a body represents a physical attribute of a body analogous to its mass. Just as the momentum of a particle is the product of its mass and velocity, in the same way the angular momentum of a rigid body rotating about an axis is the product of its moment of inertia I about that axis and its angular velocity.

12-3 Calculation of moments of inertia

Consider, in Figure 12-4, a rigid body whose moment of inertia about some axis we wish to compute. Let us set up a coordinate system and select the z-axis to coincide with the axis of interest. The perpendicular distance of a typical constituent of mass m_i with coordinates (x_i, y_i, z_i) from this axis of rotation is $(x_i^2 + y_i^2)^{1/2}$. Substitution into (12-4) yields for the moment of inertia I about the z-axis

$$I = \sum m_i (x_i^2 + y_i^2) \tag{12-5}$$

where the summation is to be carried out over all particles of the body. For continuous distributions of matter the appropriate generalization of (12-5) is

$$I = \int (x^2 + y^2)\, dm \tag{12-6}$$

where dm is an infinitesimal mass element located at the point with coordinates (x, y, z), and the integral is to be carried over all mass elements dm of the body.

Associated with any given rigid body, there are, in general, a variety of moments of inertia, depending on the particular axis of interest. Of considerable importance in this connection are the moments of inertia of a rigid body about those axes which go through its center of mass. Table 12-1 lists the moments of inertia of several homogeneous bodies obtained by use of

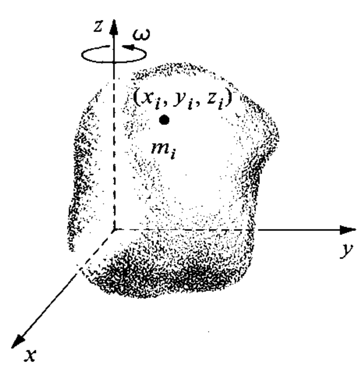
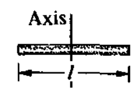
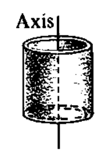
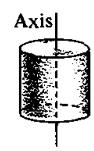
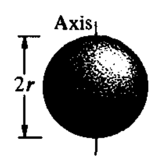
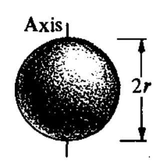
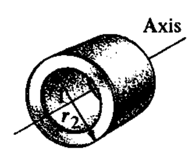
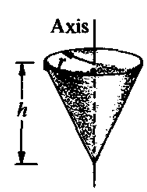
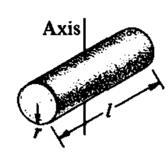


Figure 12-4

(12-6). In each case the axis is presumed to go through the center of mass and to be oriented as described in the second column. Details of the calculation of some of the entries in this table will be found in the problems and in the examples below.

There is a surprisingly simple relation between the moment of inertia I_c about an axis through the center of mass of a body and the moment of inertia of that same body about any other axis parallel to this one. This relation,

Table 12-1 Moments of inertia of homogeneous bodies (axis through the mass center)

Body	*Axis*	*Moment of inertia*	
Uniform rod of length l	Perpendicular to rod	$\frac{ml^2}{12}$	
Cylindrical shell of radius r	Axis of cylinder	mr^2	
Solid cylinder of radius r	Axis of cylinder	$\frac{mr^2}{2}$	
Solid sphere of radius r	Diameter	$\frac{2}{5}mr^2$	
Spherical shell of radius r	Diameter	$\frac{2}{3}mr^2$	
Hollow cylinder of inner radius r_1 and outer radius r_2	Axis of cylinder	$\frac{m}{2}(r_1^2 + r_2^2)$	
Right circular cone of altitude h and base radius r	Axis of cone	$\frac{3}{10}mr^2$	
Solid cylinder of length l and radius r	Central diameter	$\frac{1}{4}m\left(r^2 + \frac{l^2}{3}\right)$	

which is derived in Appendix E, is known as the parallel-axis theorem. It may be stated as follows:

If I_c is the moment of inertia of a body of mass M about an axis through its center of mass, and if I is the moment of inertia of the same body about a parallel axis at a perpendicular distance D from the center-of-mass axis, then

$$I = I_c + MD^2 \tag{12-7}$$

The meaning of the various quantities in this relation is exemplified in Figure 12-5. Here, I represents the moment of inertia of the body about the z-axis and I_c is the corresponding moment of inertia of the body about a parallel axis going through the center of mass and at a distance D from the z-axis. Evidently the distance D is related to the coordinates of the center of mass of the rigid body by $D = (X_c^2 + Y_c^2)^{1/2}$.

Because of the parallel-axis theorem in practice we need concern ourselves only with calculating the moments of inertia of a body about the axes through its center of mass. It is for this reason that I_c is such an important parameter.

Example 12-1 Two particles of masses m_1 and m_2 are at the ends of a massless rigid rod of length l. Calculate the moment of inertia I for this body about an axis perpendicular to the rod and at a distance a from m_1.

Solution Let us set up a coordinate system as in Figure 12-6, with the axis of rotation along the z-direction and with the rod along the y-axis. Since the x-coordinates of both particles vanish, it follows by use of (12-5) that

$$I = \sum m_i(0 + y_i^2) = m_1a^2 + m_2(l - a)^2$$

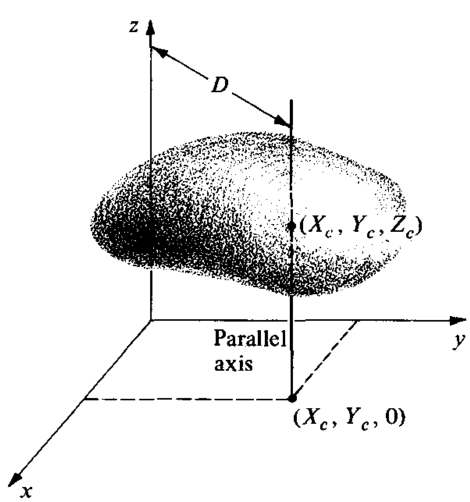


Figure 12-5

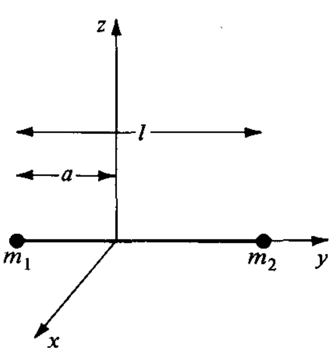


Figure 12-6

For the special case $a = lm_2/(m_1 + m_2)$ for which the z-axis goes through the center of mass, consistent with (12-7) this formula reduces to

$$I_c = \frac{m_1 m_2 l^2}{m_1 + m_2}$$

Example 12-2 Four particles of masses 1 kg, 2 kg, 3 kg, and 4 kg are the vertices of a square of side $a = 0.5$ meter. Assuming that the mass of the connecting rods can be neglected, calculate the moment of inertia of the structure about an axis perpendicular to the plane of the square and going through the intersection of its diagonals.

Solution Since each diagonal of a square of side a has length $a\sqrt{2}$, it follows that the perpendicular distance of each of the four particles from the axis is $\frac{1}{2}\sqrt{2} \times 0.5$ meter = 0.35 meter. Therefore, (12-4) yields

$$I = \sum m_i R_i^2 = (1\text{ kg} + 2\text{ kg} + 3\text{ kg} + 4\text{ kg}) \times (0.35\text{ m})^2$$
$$= 1.2\text{ kg-m}^2$$

Example 12-3 Calculate the moment of inertia of a thin, uniform rod of mass m and length l about an axis perpendicular to the rod and at a distance a from one end.

Solution Let us set up, as in Figure 12-7, a coordinate system with the z-axis along the axis of rotation and with the rod along the y-axis. Consider an element of the rod of length dy at a distance y from the z-axis. Since the rod is uniform, the mass dm of this element is

$$dm = \frac{m}{l}\,dy$$

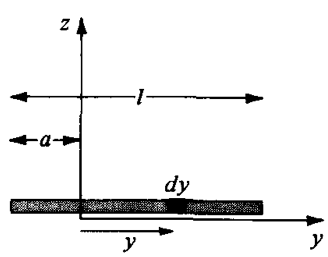


Figure 12-7

Substituting into (12-6), we find that

$$I = \int (0 + y^2)\,dm = \int_{-a}^{l-a} \frac{m}{l} y^2\,dy = \frac{m}{l}\frac{y^3}{3}\bigg|_{-a}^{l-a}$$
$$= \frac{m}{3l}[(l-a)^3 + a^3]$$

where the first equality follows since the x-coordinate of each element of the rod vanishes.

The special cases $a = 0$ and $a = l/2$ correspond, respectively, to rotations about one end and about the center. Here

$$I = \frac{ml^2}{3} \qquad \text{(axis through one end)}$$

$$I_c = \frac{ml^2}{12} \qquad \text{(axis through center of mass)}$$

where the symbol I_c has been introduced in the second formula since the center of mass of a uniform rod is at its geometric center. By making the choice $D = \frac{1}{2}l$ it is also possible to derive one of these formulas from the other by use of the parallel-axis theorem.

Example 12-4 Calculate the moment of inertia of a very thin, cylindrical shell of radius R, length l, and mass m, about the symmetry axis of the cylinder (see Figure 12-8).

Solution Since every element of mass of a thin cylindrical shell is at the same perpendicular distance R from the axis it follows, by use of (12-4), that for this case

$$I = mR^2$$

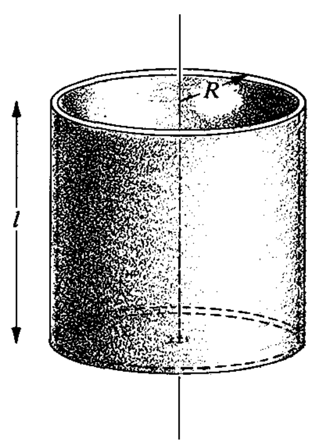


Figure 12-8

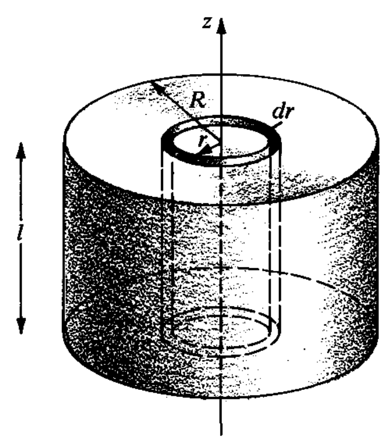


Figure 12-9

Example 12-5 Calculate the moment of inertia of a solid and homogeneous cylinder of total mass M of radius R and of length l about the symmetry axis of the cylinder.

Solution Let us set up, as in Figure 12-9, a coordinate system with the z-axis along the symmetry axis. By hypothesis, the cylinder is uniform and let $\rho_0 (= M/\pi R^2 l)$ represent the uniform mass per unit volume, or mass density, of the cylinder. According to Example 12-4, the moment of inertia dI of a thin, cylindrical shell of height l of radius r and of thickness dr is

$$dI = r^2\, dm = r^2(2\pi r\, dr\, l\rho_0)$$

since the volume of this cylindrical shell is $2\pi r\, dr\, l$. To obtain the moment of inertia of

the entire cylinder, we integrate over all values of r from 0 to R:

$$I = \int_0^R r^2\, 2\pi l \rho_0 r\, dr = 2\pi l \rho_0 \int_0^R r^3\, dr = 2\pi l \rho_0 \frac{r^4}{4}\bigg|_0^R$$
$$= \frac{\pi l \rho_0 R^4}{2}$$

where in the third equality the constant $2\pi\rho_0 l$ has been taken out from under the integral sign. Finally, since $\rho_0 = M/\pi R^2 l$, this becomes

$$I = \frac{MR^2}{2}$$

12-4 Rotation about a fixed axis

We now begin our main task of analyzing the motion of a rigid body rotating about an axis. In this section, only the special case of an axis permanently fixed in space will be considered. A cylinder rotating about its axis on a fixed shaft is an example of this type. In Section 12-7 we shall turn to the somewhat more complex case in which the axis of rotation undergoes translations parallel to itself.

Consider first a rigid body which is constrained to rotate about an axis fixed in space. Figure 12-10 shows a typical situation of this type where a cylinder of mass M rotates about the fixed shaft AA'. If $\theta \equiv \theta(t)$ is the angular displacement of a point of the body at any time t—for example, the displacement of the point P in the figure from some reference line BB'—then, according to (5-17), the angular velocity ω of the body is

$$\omega = \frac{d\theta}{dt} \tag{12-8}$$

Moreover, since each point of the body rotates at the same angular velocity ω, the displacement $\theta(t)$ of any given point also describes the angular displacement *of the body* as a whole.

Consider now a body constrained to rotate about a fixed axis under the action of various external forces and torques. Let us select an origin at some point on the axis of rotation of the body and let τ be the component, along this

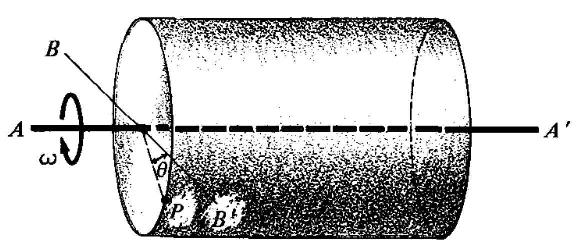


Figure 12-10

axis, of the torque produced by any such external forces. Applying Theorem II with this origin, we find on taking the component of (11-17) along the axis of rotation.

$$\tau = \frac{dL}{dt} = \frac{d}{dt}(I\omega)$$

where L is the component of the angular momentum along the axis, and where the second equality follows by use of (12-3). The symbol I here represents the moment of inertia about the given axis. Since I is constant in time, this may be expressed in the equivalent form

$$\tau = I\frac{d\omega}{dt} \tag{12-9}$$

and this is the key dynamical relation characterizing a body rotating about a fixed axis.

According to (5-17), the symbol α is also used to represent the angular acceleration $d\omega/dt$ of the body. Thus (12-9) is often written in the equivalent form

$$\tau = I\alpha \tag{12-10}$$

It is interesting to note the formal similarity between (12-9) and the one-dimensional form of Newton's law, $F = m\, dv/dt$, with τ playing the role of F, the moment of inertia I playing the role of mass m, and with the angular acceleration $\alpha = d\omega/dt$ playing the role of the linear acceleration $a = dv/dt$. In solving problems of rigid-body rotation, this formal analogy is very useful to keep in mind. The problem of finding the position $x = x(t)$ of a particle subject to, say, a constant force F_0, for example, is mathematically equivalent to the corresponding problem of finding the angular displacement $\theta = \theta(t)$ of a body subject to a constant torque τ while rotating about a fixed axis.

Example 12-6 A body of moment of inertia $I = 0.8$ kg-m^2 rotates about a fixed axis with a constant angular velocity of 100 rad/s. Calculate its angular momentum and the torque required to sustain this motion.

Solution Substituting the given data into (12-3), we obtain

$$L = I\omega = 0.8\ \text{kg-m}^2 \times 100\ \text{rad/s} = 80\ \text{J-s}$$

Since the angular velocity is constant, the angular acceleration $\alpha = d\omega/dt$ vanishes, and therefore no torque is required to sustain this motion.

Example 12-7 Consider the rotating cylinder in Figure 12-11, and suppose that $M = 5$ kg, $F_0 = 0.6$ newtons, and $a = 0.2$ meter. Calculate:

(a) The torque τ acting on the cylinder.
(b) The angular acceleration α of the cylinder.

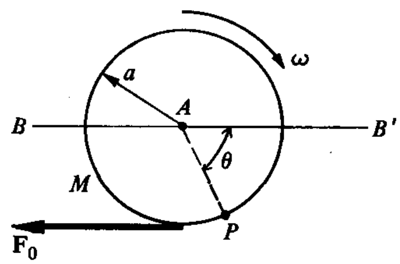


Figure 12-11

Solution

(a) The direction of the torque produced by the force $\mathbf{F}_0$ in Figure 12-11 is perpendicular to and down into the plane of the diagram. Its magnitude is

$$\tau = aF_0 = 0.2\ \text{m} \times 0.6\ \text{N} = 0.12\ \text{N-m}$$

(b) According to Table 12-1, the moment of inertia I of a cylinder about its axis is $Ma^2/2$. Hence

$$I = \frac{Ma^2}{2} = \frac{5\ \text{kg} \times (0.2\ \text{m})^2}{2}$$
$$= 0.1\ \text{kg-m}^2$$

and thus, according to (12-10),

$$\alpha = \frac{\tau}{I} = \frac{0.12\ \text{N-m}}{0.1\ \text{kg-m}^2} = 1.2\ \text{rad/s}^2$$

12-5 Work and kinetic energy

The purpose of this section is to explore in more detail the formal analogy between (12-9) and the one-dimensional form of Newton's second law and to derive in this connection the work-energy theorem for rigid bodies.

Let us first calculate the kinetic energy T of a rigid body rotating at the instantaneous angular velocity ω about a fixed axis. Consider, in Figure 12-3, a typical particle of mass m_i at a perpendicular distance R_i from the axis of rotation of a rigid body. The kinetic energy T_i of this particle is

$$T_i = \frac{1}{2} m_i v_i^2 = \frac{1}{2} m_i R_i^2 \omega^2$$

and the total kinetic energy T is obtained by summing over all particles

$$T = \sum T_i = \frac{1}{2} \omega^2 \sum m_i R_i^2$$

where the second equality follows since the angular velocity ω of each particle of a rigid body is the same. According to (12-4), this may be expressed

equivalently as

$$T = \frac{1}{2} I\omega^2 \tag{12-11}$$

where I is the moment of inertia of the body about the given axis.

In order to relate the kinetic energy of a rotating body to the work carried out on it by a torque τ, consider again the body in Figure 12-3 and suppose a force **F** to be acting at the position of the ith particle. If F is the component of this force tangent to the circular orbit of the particle, then the work dW carried out by **F** as the body rotates through a small angle $d\theta$ is

$$dW = FR_i\, d\theta$$

since the particle undergoes the displacement $R_i\, d\theta$ in this process. But FR_i is the component, τ, of the torque produced by **F** along the axis of rotation. Therefore the work dW carried out on a body by a torque τ as it is rotated through a small angle $d\theta$ is

$$dW = \tau\, d\theta \tag{12-12}$$

and the work W carried out as the body is rotated from an angle θ_1 to θ_2 is found by integration to be

$$W = \int_{\theta_1}^{\theta_2} \tau\, d\theta \tag{12-13}$$

Although derived only for the special case of a single force, (12-12) and (12-13) can be shown to be very generally applicable.

Suppose now a rigid body rotating about a fixed axis under the action of a torque τ. The change dT in its kinetic energy during a time interval dt may be expressed as

$$dT = \frac{dT}{dt}\, dt = \frac{d}{dt}\left(\frac{1}{2} I\omega^2\right) dt = I\omega \frac{d\omega}{dt}\, dt = \tau\omega\, dt$$

where the final equality follows by use of (12-9). But according to (12-8), $\omega\, dt = d\theta$. Hence, making use of (12-12), this may be expressed as

$$dT = dW$$

and this integrates to

$$W = \frac{1}{2} I\omega_2^2 - \frac{1}{2} I\omega_1^2 \tag{12-14}$$

where W is the work carried out by the torque while the body goes from an initial angular velocity ω_1 to a final one ω_2. The similarity between (12-14) and (7-22), as well as that between (12-11) and (7-23), should be noted.

This formal analogy between the above formulas and the corresponding ones for the one-dimensional motion of a particle is now complete. Table 12-2 presents a listing that spells out this one-to-one correspondence in detail. According to the table, for example, the analogue of ω is v, and that of m is I.

Table 12-2

One-dimensional motion of a particle		*Rotation of a rigid body about a fixed axis*	
Physical quantity	Symbol	Physical quantity	Symbol
Position	x	Angular position	θ
Velocity	$v = dx/dt$	Angular velocity	$\omega = d\theta/dt$
Acceleration	$a = dv/dt$	Angular acceleration	$\alpha = d\omega/dt$
Force	F	Torque	τ
Mass	m	Moment of inertia	I
Momentum	mv	Angular momentum	$I\omega$
Kinetic energy	$\frac{1}{2}mv^2$	Rotational kinetic energy	$\frac{1}{2}I\omega^2$
Work	$F\,dx$	Work	$\tau\,d\theta$

Therefore, consistent with this table we find that the analogue of the kinetic energy of a particle, $\frac{1}{2}mv^2$, is $\frac{1}{2}I\omega^2$. Moreover, since the analogue of the angular displacement θ is the position x, it follows from the definition in Chapter 8 that a torque τ will be conservative, provided that there exists a potential function $V \equiv V(\theta)$ so that the work W carried out by τ when the body undergoes an angular displacement $(\theta_2 - \theta_1)$ is the difference $[V(\theta_1) - V(\theta_2)]$. Without further ado, then, it follows that the quantity $(\frac{1}{2}I\omega^2 + V(\theta))$ is a constant of the motion in this case. This correspondence between these two types of physical quantities, it should be noted, is of a purely formal nature and does not imply that these physical systems are similar in other respects. Thus, Table 12-2 should be viewed simply as a mnemonic device which can be used to obtain relations involving quantities in the right-hand column in terms of a knowledge of relations involving those on the left.

Example 12-8 A homogeneous cylinder of radius $a = 40$ cm and of mass $M = 10$ kg is free to rotate about its axis on frictionless bearings. Suppose, as shown in Figure 12-11, that a force $\mathbf{F}_0$ of magnitude 10 newtons is suddenly applied at the rim.

(a) What is the angular acceleration α of the wheel?

(b) What is the angular velocity ω and the kinetic energy of the wheel at time $t = 2$ seconds?

(c) How much work does the force carry out on the wheel during this 2-second interval?

Solution The moment of inertia I of the cylinder about its axis is, according to Table 12-1,

$$I = \frac{Ma^2}{2} = \frac{1}{2} \times 10\text{ kg} \times (0.4\text{ m})^2 = 0.8\text{ kg-m}^2$$

(a) Solving (12-10) for the angular acceleration α we obtain

$$\alpha = \frac{\tau}{I} = \frac{10\text{ N} \times 0.4\text{ m}}{0.8\text{ kg-m}^2} = 5\text{ rad/s}^2$$

(b) Integrating the relation

$$\alpha = 5\ \text{rad/s}^2 = \frac{d\omega}{dt}$$

we find by using the initial conditions $\omega(0) = 0$ that

$$\omega(t) = (5\ \text{rad/s})t$$

At $t = 2$ seconds this has the value $\omega(2s) = 10$ rad/s and corresponds to the kinetic energy T

$$T = \frac{1}{2}I\omega^2 = \frac{1}{2} \times 0.8\ \text{kg-m}^2 \times (10\ \text{rad/s})^2$$
$$= 40\ \text{J}$$

(c) Since the wheel starts from rest, it follows that the increase in its kinetic energy during this time interval is 40 joules. According to the work-energy theorem, this must be equal to the work carried out by the external torque on the cylinder.

Equivalently, according to (b), the wheel turns through an angle of 10 rad during this time interval and, since the torque τ acting on it has the constant value of 4.0 newton meters, we have

$$W = \int \tau\, d\theta = \tau\theta = (4\text{N-m}) \times 10\ \text{rad} = 40\ \text{J}$$

Example 12-9 A 40-kg homogeneous sphere of radius 10 cm is at a certain instant rotating about a shaft through its center at 600 rpm. Assuming that a constant frictional torque acts so that the sphere comes to rest in 10 seconds, calculate the magnitude of this torque.

Solution The situation is shown in Figure 12-12. First, according to Table 12-1, the moment of inertia of this sphere about an axis through its center is

$$I = \frac{2}{5}ma^2 = 0.4 \times 40\ \text{kg} \times (0.1\ \text{m})^2 = 0.16\ \text{kg-m}^2$$

The initial angular velocity ω_0 of the sphere may be computed in the following way. The quantity 600 rpm corresponds to 600 revolutions per minute, or 10 revolutions per second. Since there are 2π radians per revolution, it follows that $\omega_0 = 2\pi \times 10$ rad/s $= 20\pi$ rad/s.

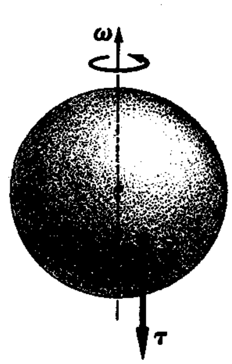


Figure 12-12

Now if $-\alpha_0$ ($\alpha_0 > 0$) is the presumed constant angular acceleration of the wheel, we find by integrating the relation $-\alpha_0 = d\omega/dt$ that

$$\omega(t) = -\alpha_0 t + \omega_0$$

with $\omega_0 = 20\pi$ rad/s, the initial angular velocity of the sphere. Therefore, since the sphere comes to rest at time $t = 10$ seconds, it follows that $\omega(10\text{ s}) = 0$. Setting $t = 10$ seconds and $\omega = 0$, and solving for α_0, we obtain

$$\alpha_0 = \frac{\omega_0}{10\text{ s}} = \frac{20\pi\text{ rad/s}}{10\text{ s}} = 2\pi\text{ rad/s}^2$$

Substituting this value for α_0 and $I = 0.16\text{ kg-m}^2$ into (12-10), we find that

$$\tau = I\alpha = 0.16\text{ kg-m}^2 \times 2\pi\text{ rad/s}^2 = 1.0\text{ N-m}$$

In other words, a frictional torque of 1.0 N-m will cause the sphere to come to rest from its initial angular velocity of 600 rpm in 10 seconds.

12-6 The physical pendulum

A *physical pendulum* is a plane lamina which is suspended from a point in a way so that it is free to rotate about a horizontal axis perpendicular to the plane of the lamina through this point.

Consider the physical pendulum in Figure 12-13, and suppose that the horizontal axis about which it rotates goes through the point A at a distance D from the mass center at B. The position at any time t of this lamina may be described in terms of the angle θ that the line joining the center of mass at B and the point of suspension at A makes with the vertical. Only two external forces act on the pendulum. These are the weight $M\mathbf{g}$, which acts at B and is directed vertically downward, and the reaction force $\mathbf{R}$, which acts at the point of suspension. Since the point A is fixed as the lamina oscillates back and forth, we may apply Theorem II with the point A as an origin. If we assume that the contact at A is smooth, it follows that the unknown reaction force $\mathbf{R}$ produces no torque on the lamina. The total external torque $\boldsymbol{\tau}$ acting

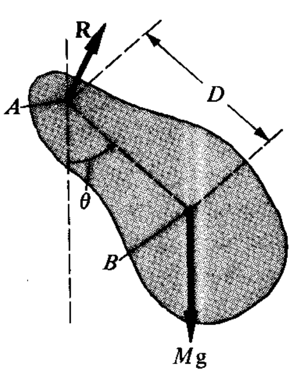


Figure 12-13

on the lamina about this origin is then due to its weight, and this is directed perpendicular to and down into the plane of the page in the figure. It has the magnitude

$$\tau = MgD \sin \theta \tag{12-15}$$

Let us adopt the convention that the angular velocity $\omega = d\theta/dt$ is positive when the pendulum is moving toward the right. Then when $\omega > 0$ the angular momentum **L** of the lamina about the point A is directed perpendicular to and up out of the plane of the figure. According to (12-7), **L** has magnitude $(I_c + MD^2)\omega$, where I_c is the moment of inertia of the lamina about an axis perpendicular to the plane and going through its mass center. Taking into account the difference of direction between the torque and the angular momentum we find, by use of Theorem II, that

$$-MgD \sin \theta = (I_c + MD^2)\frac{d\omega}{dt} \tag{12-16}$$

Equivalently, since $d\omega/dt = d^2\theta/dt^2$, this may be expressed as

$$\frac{d^2\theta}{dt^2} + \frac{g}{D}\frac{1}{[1 + I_c/MD^2]} \sin \theta = 0 \tag{12-17}$$

and the solution of this equation represents the angular position $\theta = \theta(t)$ of the lamina at any time t.

It is instructive to compare the equation of motion for the physical pendulum in (12-17) with the corresponding one for a simple pendulum in (6-28). Such a comparison shows that the motion of a physical pendulum is precisely the same as that of a simple pendulum of length l given by

$$l = D\left[1 + \frac{I_c}{MD^2}\right] \tag{12-18}$$

With this identification, all results obtained for the simple pendulum are applicable to the physical pendulum. For example, making use of (6-31), we find that the period P of a physical pendulum for small-amplitude oscillations is

$$P = 2\pi\left[\frac{D(1 + I_c/MD^2)}{g}\right]^{1/2} \tag{12-19}$$

and that its energy E, defined by

$$E = \frac{1}{2}(I_c + MD^2)\omega^2 + MgD(1 - \cos \theta) \tag{12-20}$$

is a constant of the motion.

Example 12-10 A uniform rod of mass m and length l is suspended at one end and allowed to oscillate freely in a horizontal plane (see Figure 12-14a).

(a) What is the period of this physical pendulum for small amplitudes?

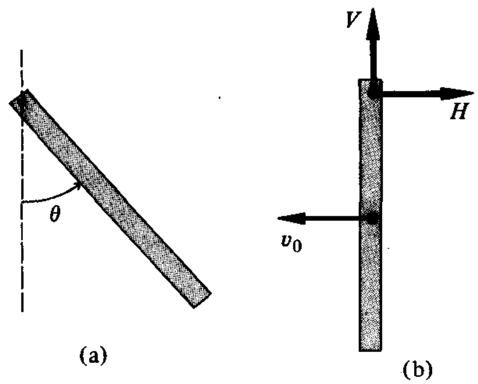


Figure 12-14

(b) If the pendulum is initially released at rest in a horizontal position (this corresponds to $\theta = 90°$ in the figure), what is the velocity v_0 of the mass center at a subsequent instant when the rod is vertical?

(c) Calculate the vertical V and the horizontal H components of the force on the rod at the point of suspension (see Figure 12-14b).

Solution

(a) According to Table 12-1, $I_c = ml^2/12$. Substituting this formula and the value $D = l/2$ into (12-19), we find the period

$$P = 2\pi\left[\frac{2l}{3g}\right]^{1/2}$$

(b) The substitution of the initial data, $\omega = 0$ and $\theta = 90°$, into (12-20) shows that $E = mgl/2$. Therefore (12-20) may be expressed in the form

$$\frac{mgl}{2} = \frac{1}{2}\left(\frac{ml^2}{3}\right)\omega^2 + \frac{1}{2}\,mgl\,(1 - \cos\theta)$$

and for $\theta = 0$ this yields $\omega = \sqrt{3g/l}$. Hence the velocity v_0 of the center of mass is

$$v_0 = \frac{1}{2}\,l\left(\frac{3g}{l}\right)^{1/2}$$

(c) According to (12-17) at $\theta = 0$, the angular acceleration $d^2\theta/dt^2$ of the rod also vanishes. Accordingly, it follows that at the instant shown in Figure 12-14b, the acceleration of the center of mass of the rod has a zero component along the horizontal direction, and thus $H = 0$.

With regard to the vertical component V, note that the center of mass of the rod has a centripetal acceleration, directed vertically upward. According to Theorem I, then,

$$V - mg = m\frac{v_0^2}{l/2}$$

and substituting for v_0 by use of the result of (b), we obtain

$$V = \frac{5}{2}\,mg$$

12-7 Motions involving both translations and rotations

So far only the possibility of the rotation of the body about an axis fixed in space has been considered. The purpose of this section is to study the slightly more complex case for which the axis of rotation also accelerates although its orientation continues to be fixed. Because of this acceleration it is not generally possible to apply Theorem II by selecting an origin on the axis of rotation, as we did above. Therefore, let us confine the following discussion to those physical situations for which the center of mass lies on the axis of rotation, so that Theorem II is still applicable.

One of the important tools that can be used to analyze motion of this type is the conservation-of-energy law. Of particular interest in this connection is the case of a cylinder or of a sphere rolling along a flat or curved surface. For this case we find that the energy of the body is a constant of the motion, provided that there is no slipping.

In order to establish this let us first define the term "no slipping" more precisely. Consider, in Figure 12-15a, a disk of radius a rotating at the angular velocity ω about a fixed axis perpendicular to the plane of the disk at its center C. As shown, the top point of the wheel, point A in the figure, has the velocity $a\omega$ directed to the right, while the bottom point B has a velocity $a\omega$ directed to the left. Figure 12-15b shows the same rotating wheel, but now as seen by an observer moving to the left at the velocity $a\omega$. For this observer, the point A moves to the right at the velocity $2a\omega$, the center of the wheel C travels to the right with the velocity $a\omega$, and the bottom of the wheel, at point B, is at rest. Consider now, in Figure 12-15c, an identical disk, which rolls *without slipping* on a horizontal surface, so that it has the same angular velocity ω as above about its axis. Since the disk rolls without slipping, it follows that the velocity of point C is directed to the right and has the magnitude $a\omega$. For as the center C of the disk travels a distance x to the right, since no slipping takes place, the angle θ through which the disk must rotate during this interval is $\theta = x/a$. Therefore, the velocity $v = dx/dt$ of the center of the disk is $a\,d\theta/dt = a\omega$. In addition to this translatory motion, each point of the disk also rotates at the angular velocity ω about a perpendicular axis through C. Therefore, the motion of each point of the rolling disk must be the same as that of the disk rotating about a fixed axis but as seen by the moving

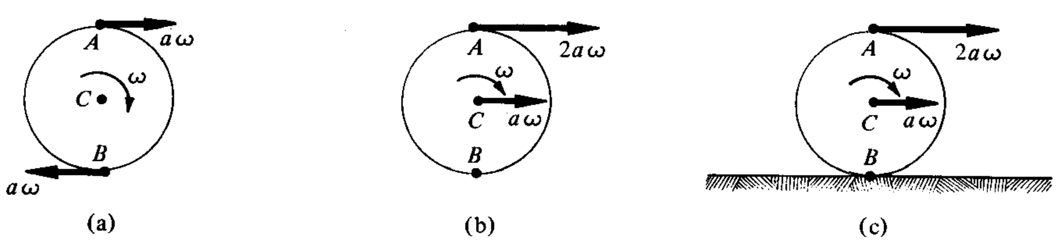


Figure 12-15

observer in Figure 12-15b. Most important, from our present point of view, is the fact that if no slipping occurs, then *the point of the disk that is instantaneously in contact with the horizontal surface* (point B in the figure) *has zero velocity*. Hence, and this is a most important point, whatever the nature of the contact force acting on a body which rolls without slipping, this force carries out *no work* on the body. Therefore, if the remaining forces acting on the body are conservative, its energy will be conserved.

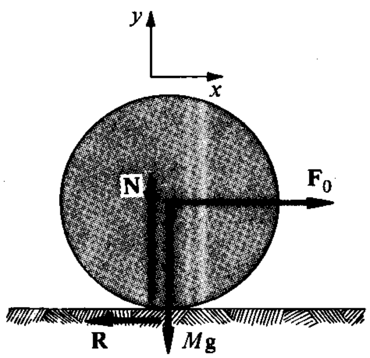


Figure 12-16

To illustrate the principle consider, in Figure 12-16, a disk of mass M and radius a forced to roll without slipping along a horizontal surface by a horizontal force $\mathbf{F}_0$ acting at its center. In addition to $\mathbf{F}_0$, there is the force of gravity $M\mathbf{g}$ and a contact force that the floor exerts on the disk. In the figure this contact force has been resolved into a normal component $\mathbf{N}$ and a horizontal component $\mathbf{R}$. Since there is no slipping, the contact force on the disk does no work on it as it rolls along. According to (8-2), the constant force $\mathbf{F}_0$ is derivable from the potential function $V = -F_0 x$, where x is the horizontal coordinate of the center of mass. Therefore, the energy E of the disk is

$$E = \frac{1}{2} I_c \omega^2 + \frac{1}{2} M(a\omega)^2 - F_0 x \tag{12-21}$$

(where ω is the angular velocity of the disk about its center and $I_c = Ma^2/2$ is the moment of inertia about the mass center) and is a constant of the motion. In writing down this relation we have used the fact (established in the problems) that the kinetic energy of a rigid body is the sum of the kinetic energies associated with the motion of the mass center $\frac{1}{2}M\mathbf{V}_c^2$, and with the motion relative to the mass center, $\frac{1}{2}I_c\omega^2$.

Alternatively, we may apply Theorems I and II directly. Since $a\omega$ is the horizontal velocity of the center of mass of the cylinder, Theorem I implies that

$$F_0 - R = M\frac{d}{dt}(a\omega) \tag{12-22}$$

Moreover, since the only external force that produces a torque about the center of mass is **R**, it follows by use of Theorem II that

$$Ra = I_c \frac{d\omega}{dt} \tag{12-23}$$

If (12-22) is multiplied by a and added to (12-23), the time derivative of (12-21) divided by ω results. This confirms the fact that if no slipping occurs, then the contact force does no work and the energy of the disk is conserved.

Example 12-11 A spool of thread of radius a and mass M has the shape of a uniform cylinder. Suppose that one end of the thread is attached to a ceiling and the spool is allowed to unwind under the action of gravity (see Figure 12-17). Neglecting the mass and thickness of the thread, calculate the acceleration of the spool and the tension T in the thread.

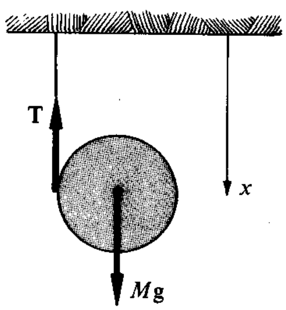


Figure 12-17

Solution The only external forces acting on the spool are its weight $M\mathbf{g}$ and the upward force of the tension **T** in the string. If x is the distance of the mass center below the ceiling at time t and $v = dx/dt$ is the downward velocity of the mass center, then assuming that the string does not stretch, it follows that $v = a\omega$, where ω is the instantaneous angular velocity of the spool about its center. According to Theorem I,

$$Mg - T = M\frac{dv}{dt}$$

and, applying Theorem II with the origin at the mass center, we obtain

$$aT = \frac{Ma^2}{2}\frac{d\omega}{dt}$$

since the moment of inertia I_c of a homogeneous cylinder about the center of mass is $Ma^2/2$. These two relations and the "no slipping" condition $v = a\omega$ constitute a complete set of working equations. Their solution determines the values for the three unknowns T, dv/dt, and $d\omega/dt$ to be

$$T = \frac{1}{3}Mg \qquad \frac{dv}{dt} = \frac{2}{3}g \qquad \frac{d\omega}{dt} = \frac{2}{3}\frac{g}{a}$$

An alternate way of obtaining the acceleration of the center of mass is by use of the conservation-of-energy law. Since the spool rolls without sliding along the unwinding

string, it follows that the tension T in the string does no work. Accordingly, the energy E, defined by

$$E = \frac{1}{2} Mv^2 + \frac{1}{2}\left(\frac{Ma^2}{2}\right)\left(\frac{v}{a}\right)^2 - Mgx$$

is conserved. Differentiating this relation and making use of the fact that $v = dx/dt$, we find, as above, that the downward acceleration is $2g/3$.

Example 12-12 Analyze the motion of a homogeneous sphere of mass M and radius a, which rolls without slipping down a fixed inclined plane of angle β.

Solution As shown in Figure 12-18, there are two forces acting on the sphere: the force of gravity $M\mathbf{g}$, which acts at the center of the sphere, and a contact force, which is conveniently resolved into a normal component N and a tangential component R. Let us define a coordinate system so that x represents the position of the center of mass along the inclined plane. The velocity of the center of the sphere down the inclined plane is then $v = dx/dt$ and, since in addition the sphere rolls without slipping, it follows that its angular velocity ω about the mass center is $\omega = v/a$.

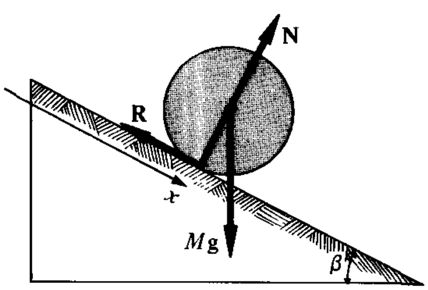


Figure 12-18

According to Theorem I,

$$Mg \sin\beta - R = M\frac{dv}{dt} \tag{12-24}$$

and applying Theorem II about the center of the sphere we obtain

$$Ra = \frac{2}{5} Ma^2 \frac{d\omega}{dt} \tag{12-25}$$

The combination of (12-24) and (12-25) and the condition of no slipping, $v = a\omega$, constitutes a complete description for the motion of the sphere. For example, eliminating R and ω we find that

$$\frac{dv}{dt} = \frac{5}{7} g \sin\beta$$

It is left as an exercise to confirm that the time derivative of the energy

$$E = \frac{1}{2} Mv^2 + \frac{1}{2}\left(\frac{2}{5} Ma^2\right)\omega^2 - Mgx \sin\beta$$

yields the same formula for the acceleration as above.

12-8 Conservation of angular momentum

Consider a body rotating about a fixed axis. If no external torques act on it, then its angular momentum $L = I\omega$ is a constant of the motion. For situations studied up to this point, the moment of inertia I of the body has been always assumed to be constant. The fact that the angular momentum of the body is conserved implies for these cases that its angular velocity ω is also constant in time.

However, physical situations also arise for which a body rotates about a fixed axis, and even though it is not subject to any external torques its angular velocity ω still changes. Since the angular momentum of a body is given by the product $I\omega$, it follows that if ω varies in time then so must I, and in such a way that their product remains constant. The purpose of this section is to illustrate some experiments that exemplify this possibility.

Figure 12-19a shows a man standing at the center of a horizontal platform, which rotates on frictionless bearings at the constant angular velocity $\boldsymbol{\omega}_0$ about a vertical axis through its center. Suppose that the man holds in each of his outstretched hands a heavy dumbbell, and let I_0 be the total moment of inertia of the man, the dumbbells, and the platform. Since there are no external torques acting on this total system, its angular momentum $I_0\omega_0$ about the vertical axis is constant in time.

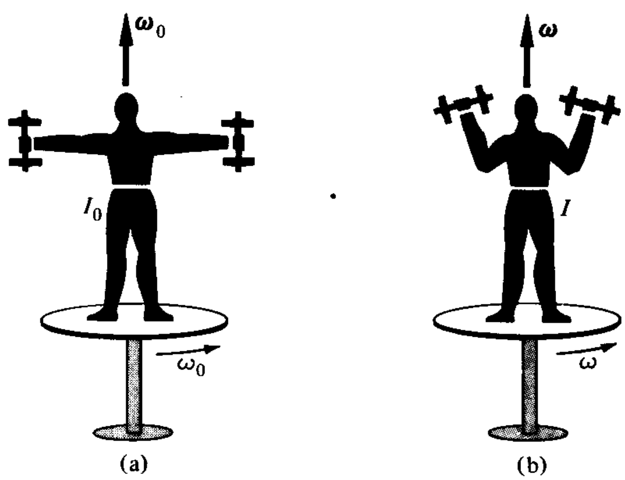


Figure 12-19

Suppose now that the man suddenly pulls in his outstretched arms, so that the final situation is as shown in Figure 12-19b. Again, since no *external* torques have acted on the system before, after, or during the change in the position of the dumbbells, it follows that its angular momentum is conserved. Therefore, if I is the moment of inertia of the man, the platform, and the dumbbells in the final configuration, his angular velocity ω must be given by

$$I_0\omega_0 = I\omega \qquad \textbf{(12-26)}$$

According to the definition in (12-4), $I < I_0$. Hence we conclude that $\omega > \omega_0$. In other words, and this is an easily observed effect, as the man pulls in the dumbbells his angular velocity increases! Correspondingly, if he subsequently extends his arms again, his angular velocity will decrease.

Figure 12-20 shows another application of this principle. Figure 12-20a shows a man standing at the center of a horizontal platform, which is again mounted on frictionless bearings and is originally stationary. In his hands he holds a shaft, about which rotates a disk at some angular velocity ω_0 so that the angular momentum of the entire system $\mathbf{L}_0$ is directed vertically upward.

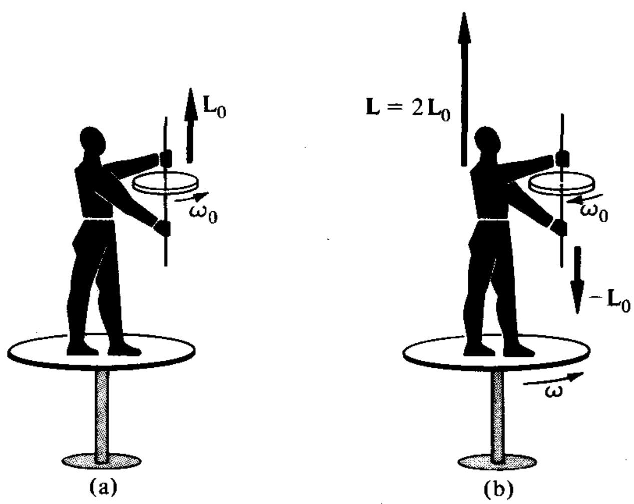


Figure 12-20

Suppose that he suddenly turns the shaft through 180° so that the angular momentum $\mathbf{L}_0$ of the disk becomes directed vertically downward. The final situation is shown in Figure 12-20b. Apparently, the angular momentum of the disk has undergone a change $-2\mathbf{L}_0$. Since there are no *external* torques acting on the total system, consisting of the disk, its shaft, the man, and the platform, it follows that in order to conserve angular momentum the platform itself must begin to rotate at some angular velocity ω. In other words, since the total angular momentum of this system must be $\mathbf{L}_0$, and since the angular momentum of the disk alone is finally $-\mathbf{L}_0$, it follows that the platform must rotate in such a way as to make up the difference $+2\mathbf{L}_0$. It is very instructive to perform this experiment for oneself in the laboratory. Physically, you will find that in order to turn the shaft of the rotating disk, you must push against the surface of the platform in such a way as to cause it to rotate in the direction shown in the figure.

Example 12-13 Suppose a star with a mass M $(= 2.0 \times 10^{30}$ kg) and radius R_s $(= 7.0 \times 10^5$ km) of our sun collapsed to the size of a *white dwarf*, with a radius $R = 10^4$ km. If the original period of rotation of the star about its axis is $P_s = 3.0 \times 10^6$ second ($\cong 1$ month), what is the period P of the white dwarf? Neglect any mass ejected during the collapse.

Solution Since for a sphere $I_c = 2MR^2/5$ and since ω and the period of rotation P are related by $\omega = 2\pi/P$, it follows from (12-26) that

$$\frac{2}{5}MR_s^2\frac{2\pi}{P_s} = \frac{2}{5}MR^2\frac{2\pi}{P}$$

Solving for P and substituting the given data, we find that

$$P = P_s\left(\frac{R}{R_s}\right)^2 = 3.0\times 10^6\ \mathrm{s}\left(\frac{10^4\ \mathrm{km}}{7.0\times 10^5\ \mathrm{km}}\right)^2$$
$$\cong 600\ \mathrm{s}$$

which is a very short period indeed considering the massive nature of a white dwarf.

†12-9 The gyroscope

One of the very interesting and unusual physical systems that can be described in detail by use of Theorems I and II is a device known as a *gyroscope.* Because of the fact that an essential feature of the motion of a gyroscope involves a change in the direction of the axis of rotation, it is not possible to analyze this problem by the methods of the preceding sections. Therefore, let us simply describe the phenomenon and attempt to understand it qualitatively.

Figure 12-21 shows a very massive cylindrical disk, of mass M and radius a, which is free to rotate on frictionless bearings about a very light and thin shaft along its axis. Suppose that one end of the shaft is supported by a fulcrum at point A, which is a distance l from the disk. With the shaft kept horizontal, suppose that the disk is set rotating about its axis at some angular velocity ω and then released. What is the resulting motion?

Offhand one might think that, since the only two forces acting on the system are the downward weight Mg of the disk and the reaction force $\mathbf{R}$ at the support point A, the disk would simply fall down. Indeed, if the disk were not

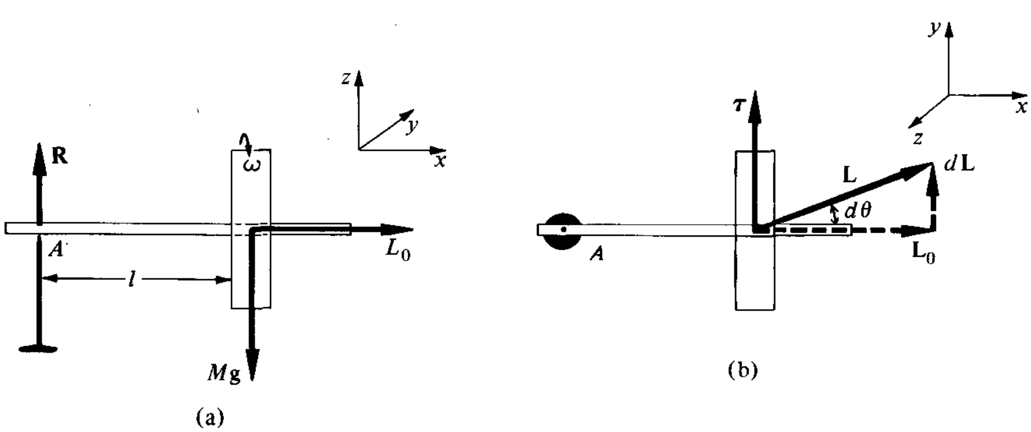


Figure 12-21

rotating so that its initial angular momentum $\mathbf{L}_0$ were zero, this is precisely what would happen. However, because the initial angular momentum of the disk does not vanish, there occurs an altogether different type of motion, which we shall now describe.

In general, the reaction force **R** is not known. Therefore, it is not possible to apply Theorem I directly in order to describe the motion of the center of mass of the disk. However, since the point A at which the reaction force **R** acts is a point fixed in space, we may apply Theorem II, with this point as an origin. The reaction force **R** produces no torque about this point.

However, the force of gravity does produce a torque about point A. The magnitude of this torque $\boldsymbol{\tau}$ is Mgl and it is directed as shown in the top view in Figure 12-21b. The change of angular momentum $d\mathbf{L}$ produced by this torque during an infinitesimal time interval dt is parallel to the torque and has the magnitude

$$dL = \tau\, dt = Mgl\, dt$$

Therefore, the angular displacement $d\theta$ of the angular-momentum vector during this time interval dt is

$$d\theta = \frac{dL}{L_0} = \frac{Mgl\, dt}{(Ma^2/2)\omega}$$

Therefore, provided that the angular velocity ω of the disk does not vanish, the disk does not fall downward under the action of gravity. Instead, the shaft rotates or *precesses* in a horizontal plane, the x-y plane in Figure 12-21, about a vertical axis through the support point A. The angular velocity $\Omega = d\theta/dt$ of this precession is

$$\Omega = \frac{\tau}{I\omega} = \frac{2gl}{\omega a^2} \tag{12-27}$$

12-10 Summary of important formulas

The component L of the angular momentum along the axis of rotation of a rigid body rotating at an angular velocity ω is

$$L = I\omega \tag{12-3}$$

The coefficient I is the *moment of inertia* of the body about the given axis and is defined by

$$I = \sum m_i R_i^2 \tag{12-4}$$

with R_i the perpendicular distance from the ith particle to the axis of rotation.

For a body constrained to rotate about a fixed axis, the component τ of the torque along the axis of rotation produces an angular acceleration $\alpha = d\omega/dt$ given by

$$\tau = I\frac{d\omega}{dt} \tag{12-9}$$

The kinetic energy T of a rotating body is

$$T = \frac{1}{2} I\omega^2 \tag{12-11}$$

and the work dW carried out on a rigid body by a torque τ in producing an angular displacement $d\theta$ is

$$dW = \tau \, d\theta \tag{12-12}$$

The total work W carried out on a rigid body as its angular velocity changes from an initial value ω_1 to a final value ω_2 is

$$W = \frac{1}{2} I\omega_2^2 - \frac{1}{2} I\omega_1^2 \tag{12-14}$$

QUESTIONS

1. Define or describe what is meant by (**a**) rigid body; (**b**) moment of inertia; (**c**) the parallel-axis theorem; (**d**) physical pendulum; and (**e**) gyroscope.
2. A rigid body rotates at the angular velocity ω about a fixed axis. Describe the characteristics of the motion of its constituent particles. What would the motion of the constituent particles be if the axis of rotation were not fixed but were itself moving, but in such a way as to remain parallel to itself?
3. For a body rotating at an angular velocity ω about an axis define what is meant by the vector $\boldsymbol{\omega}$.
4. Show, by reference to a particular example, that the angular momentum $\mathbf{L}$ of a rigid body rotating about a fixed axis need not be parallel to the angular velocity $\boldsymbol{\omega}$ of that body.
5. For each of the following pairs of bodies, state which has the larger moment of inertia about the symmetry axis through the center of the body.
   (**a**) Two cylinders with the same outer radius, the same mass, and the same length, but one of which is hollow.
   (**b**) Two spheres, with the same mass and the same outer radius, but one of which is hollow.
   (**c**) Two homogeneous cylinders of the same mass and the same outer radius, but of different lengths.
6. Consider a plane lamina in the form of an ellipse. About which axis, the semimajor or the semiminor, would you expect it to have the larger moment of inertia? Explain why its moment of inertia about an axis perpendicular to the plane of the lamina and going through its center is the same as the sum of the moments of inertia about these two axes.
7. Why are the moments of inertia of the earth about all diameters lying in the equatorial plane the same? What can be concluded from the observational fact that the moment of inertia of the earth about its actual axis of rotation differs from the corresponding moment about any diameter lying in the equatorial plane?
8. Consider a homogeneous cube. Explain why its moment of inertia about an axis perpendicular to a face of the cube and going through a center, is the same as the moment of inertia

about a diagonal. (*Note*: Detailed calculations show that the moments of inertia of a cube about *all* axes through its center are the same.)

9. Consider, in Figure 12-22, a rectangular plane lamina of mass M and sides a and b. Making use of the first entry in Table 12-1, explain why its moment of inertia about an axis AA' in the plane of the lamina and through its center is $Ma^2/12$.

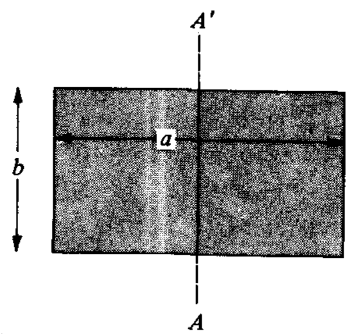


Figure 12-22

10. Making use of the result of Question 9 and (12-6), show that the moment of inertia of the rectangular slab in Figure 12-22 about an axis perpendicular to the plane of the lamina and going through its center is $M(a^2 + b^2)/12$.

11. Imagine yourself riding a bicycle due north along a horizontal road. What is the direction of the angular momentum of the bicycle relative to a fixed point on the road?

12. Suppose you are riding a bicycle due north on a horizontal road and suddenly you lean over toward your left. What is the direction of the torque on the bicycle with respect to a point on the road instantaneously in contact with one of the wheels? Explain why the bicycle will have a tendency to turn left as a result of this torque.

13. If the cylinder in Figure 12-18 is released from rest at the top of the inclined plane, and if the contact between the plane and the sphere is smooth, so that $R = 0$, what will be the angular momentum of the sphere about its center when it reaches the bottom?

14. Suppose that in Figure 12-18 a sphere rolls down an inclined plane and then onto a horizontal surface. Assume that it starts from rest, that no slipping occurs as it rolls down the inclined plane, and that the contact between the sphere and the horizontal part of the surface is perfectly smooth. **(a)** During what portions of this motion is the energy of the sphere conserved? **(b)** During what part of the journey is the velocity of the point instantaneously in contact with the surface zero?

15. Consider a football rolling, without slipping, down the side of a hill. Explain why it is that the point of the football instantaneously in contact with the hill has zero velocity.

16. Explain why the total energy of the football in Question 15 is conserved, regardless of the shape of the hill or of the football.

17. Suppose the man in Figure 12-19a suddenly drops the dumbbells. Does the angular velocity of the rotating platform change while they are falling? Will its angular velocity change in any way subsequently if they land on the platform? What if they land on the ground beyond the platform?

18. It is known today that the angular velocity of the earth about its axis is steadily decreasing, corresponding to a lengthening of our day by about 10^{-3} second each century. What is the nature of the external forces whose torque produces this change in the angular momentum of the earth?

19. Explain briefly by use of the ideas of conservation of angular momentum, how a person in the act of diving into a swimming pool can control the rate

at which he rotates about a horizontal axis through his center of mass by moving his body appropriately.

20. You are given two spheres, both of which appear to be identical in that they have the same mass and radius. Devise an experiment to determine which one of them is hollow.

PROBLEMS

1. A body of moment of inertia $I = 1.5$ kg-m^2 rotates about a fixed axis at an angular velocity $\omega = 30$ rad/s. Calculate its **(a)** angular momentum and **(b)** kinetic energy.
2. What is the kinetic energy and the angular momentum of an automobile tire of moment of inertia $I = 6$ kg-m^2 and rotating at 600 rpm about its axis?
3. Assuming that the earth is a uniform sphere of mass $M = 6.0 \times 10^{24}$ kg and of radius $R_0 = 6.4 \times 10^6$ meters, calculate its moment of inertia about a diameter.
4. **(a)** Calculate the angular momentum of the earth due to its rotation about its axis.
   (b) Compare your answer to (a) with the angular momentum of the earth due to its orbital motion (assume that it travels in a circle of radius 1.50×10^{11} meters) about the sun.
5. Calculate the angular momentum of a star of mass 2.0×10^{30} kg and radius 7.0×10^5 km if it makes one complete rotation about its axis once in 20 days. What is its kinetic energy?
6. A rigid body consists of four small weights, each of mass m and connected by four very light rods, each of length l, to form a square.
   (a) Calculate the moment of inertia of the structure about an axis perpendicular to the plane of the square and going through its center.
   (b) Calculate its moment of inertia about an axis perpendicular to the plane of the square but going through one of the weights.
   (c) Compare your result to (b) with the corresponding one obtained by use of the parallel-axis theorem and your result to (a).
7. For the structure of Problem 6, assume that $m = 100$ grams and $l = 20$ cm and calculate:
   (a) Its moment of inertia about an axis lying in the plane of the square, parallel to a side, and going through the center.
   (b) The moment of inertia about an axis along one of the diagonals of the square. Compare this with your result to (a).

*8. For the structure in Problem 6, show that the moment of inertia about an axis in the plane of the square, going through its center and making an angle α with one of the sides, is independent of α.

9. Using the notation and methods of Example 12-5, show that the moment of inertia about the symmetry axis of a hollow, homogeneous cylinder of mass M and of inner radius R_1 and outer radius R_2 is

$$I_c = \frac{1}{2} M(R_2^2 + R_1^2)$$

10. Making use of Table 12-1 and the parallel-axis theorem, determine the moments of inertia of the following bodies about the indicated axes. Let M be the mass of the body in each case.
    (a) A homogeneous cylinder of

radius R about an axis parallel to the symmetry axis of the cylinder and tangent to its surface.

(b) A homogeneous sphere of radius R about an axis tangent to the sphere.

(c) A spherical shell of radius R about an axis at a perpendicular distance $R/2$ from the center of the shell.

11. Consider, in Figure 12-23, a plane lamina which lies in the x-y plane of a certain coordinate system. Show that if I_1, I_2, and I_3 are the respective moments of inertia of the lamina about the x-, y-, and z-axes, then

$$I_3 = I_1 + I_2$$

(*Hint*: Why is the square of the distance of a constituent particle of the lamina with coordinates $(x_i, y_i, 0)$ from the z-axis the same as the sum of the squares of its distances from the x- and y-axes?)

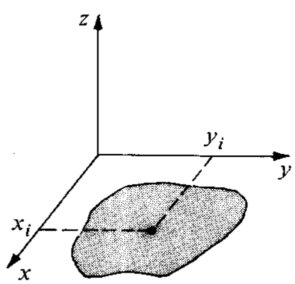


Figure 12-23

12. Show that the moment of inertia of a very thin, spherical shell of total mass M and radius R about a diameter is

$$I = \frac{2}{3} MR^2$$

Hint: Show that the moment of inertia dI of a circular ring of radius $R \sin\theta$ (as shown in Figure 12-24)

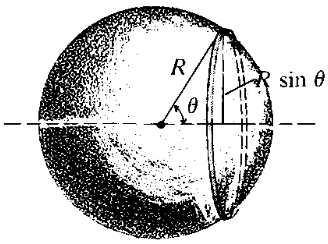


Figure 12-24

and of width $R\,d\theta$ of the spherical shell is

$$dI = \frac{M}{2} R^2 \sin^3\theta\, d\theta$$

*13. By use of the results of Problem 12, show that the moment of inertia I of a hollow, uniform sphere of inner radius R_1 and of outer radius R_2 and of mass M is

$$I = \frac{2M}{5} \frac{R_2^5 - R_1^5}{R_2^3 - R_1^3}$$

14. By use of the result of Problem 13, calculate the moment of inertia of the earth about a diameter if the earth were hollow with inner and outer radius $R_E/2$ and R_E, respectively. Assume the parameter values in Appendix B.

15. Suppose that the cylinder in Figure 12-11 has a radius of 50 cm and a mass of 5 kg, and that the tangential force F_0 has a magnitude of 15 newtons.
   (a) What is the torque on the cylinder?
   (b) What is the angular acceleration of the cylinder?
   (c) If the cylinder starts from rest, what is its angular velocity after 2 seconds?

16. Under the action of a certain constant torque, a cylinder of radius 40 cm and mass 10 kg starts from rest and attains an angular velocity about its axis of 100 rad/s in 10 seconds.

(a) What is the angular acceleration of the cylinder?
(b) What is the strength of the torque?
(c) If the torque is produced by a force acting tangentially to the cylinder, what is the strength of the force?

17. Suppose a body of moment of inertia $I = 5.0$ kg-m^2 rotates about a fixed axis at the constant angular acceleration 3 rad/s^2. Assume that initially it starts at rest.
(a) What is the angular velocity of the body at $t = 5$ seconds?
(b) What is its kinetic energy at this instant?
(c) How much work is carried out by the torque acting on the body during the first 5 seconds?

18. A spherical shell of radius 20 cm and mass 2 kg rotates about a shaft along a diameter with an angular acceleration $\alpha(t)$, given by

$$\alpha(t) = \beta t$$

where $\beta = 2$ rad/s^3. If the shell starts at rest:
(a) What is its angular velocity at any subsequent time t?
(b) Through what angular displacement will the sphere have turned at time t?
(c) What torque as a function of time is required to sustain this rate of rotation?

19. Under the action of a certain torque, a body of moment of inertia $I =$ 4 kg-m^2 rotates about a fixed axis. Suppose that at an instant when its angular velocity is 100 rad/s the torque suddenly ceases to act.
(a) What is the maximum kinetic energy of the body?
(b) What is the strength of a constant decelerating torque required so that it comes to rest 5 seconds after the original torque has ceased to act?
(c) How much work is carried out by this decelerating torque?

20. For the physical situation in Problem 15, how much work has the force $\mathbf{F}_0$ carried out on the cylinder during the 2.0-second time interval in (c)?

21. Calculate the work carried out on the cylinder by the torque in Problem 16 during the initial 10-second interval.

22. Consider, in Figure 12-25, a spool of very thin thread of total mass M and radius b supported along its axis by a smooth, fixed shaft. Suppose a block of mass m is attached to one end of the thread and allowed to fall under the action of gravity.

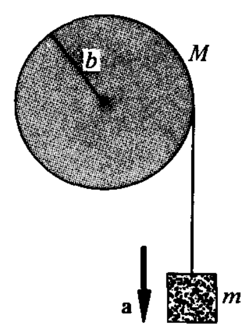


Figure 12-25

(a) Write down Newton's law for the block in terms of its downward acceleration a and the tension T in the string.
(b) Write down the torque equation for the cylinder.
(c) Show that the downward acceleration a of the block is

$$a = \frac{g}{1 + M/2m}$$

23. Consider again the situation in Figure 12-25. Assuming that the system is released at rest, make use of the formula given in (c) of Problem 22 to calculate:
(a) The distance the block has fallen at time t.

(b) The kinetic energy of the spool at time t.

(c) The kinetic energy of the block at time t.

*24. Consider again the situation described in Problem 22.

(a) Show the downward acceleration of the mass center of the block and the cylinder is

$$\frac{gm}{(M+m)(1+M/2m)}$$

(b) Making use of the result of (a) and Theorem I, calculate the upward force that the axis exerts on the spool.

(c) Explain in physical terms why, despite this force, the total energy of the system is conserved.

25. A physical pendulum consists of a very light rod of length 80 cm to whose lower end is attached a thin disk of radius 10 cm and of mass 0.5 kg (see Figure 12-26). (a) What is the period of this physical pendulum for small angles? (b) What is the length of the simple pendulum which has the same period?

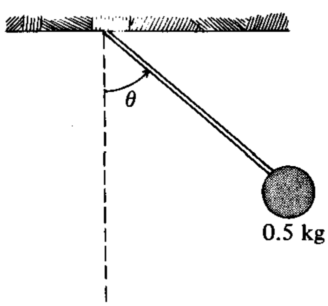


Figure 12-26

*26. Repeat Problem 25, but assume this time that the rod itself has a mass of 0.25 kg.

27. Calculate the ratio of the periods (for small angles) of the following two pendulums:

(a) A particle of mass m attached to a massless, rigid rod of length l.

(b) A small sphere of radius r and mass m attached to the same massless rod of length l.

28. Show that if a physical pendulum is released from rest at an initial angle θ_0 (see Figure 12-13), then

$$\omega^2 = \frac{2g}{D[1+I_c/MD^2]}[\cos\theta - \cos\theta_0]$$

29. Making use of the result of Problem 28 calculate at $\theta = 0$ the force $\mathbf{R}$ exerted on a physical pendulum at its point of support.

*30. Prove that the kinetic energy T of a physical pendulum may be expressed in the form

$$T = \frac{1}{2}I_c\omega^2 + \frac{1}{2}MV_c^2$$

where ω is the angular velocity of the pendulum about its center of mass and $\mathbf{V}_c$ is the velocity of the mass center. (*Hint*: Refer to Figure 12-13 and relate the magnitude of the velocity $\mathbf{V}_c$ to ω and to D.)

*31. Consider a rigid body of mass M, whose mass center has the velocity $\mathbf{V}_c$ and which rotates with an angular velocity ω about an axis through the center of mass. Generalize the result of Problem 30 by showing that the kinetic energy T of this body is

$$T = \frac{1}{2}I_c\omega^2 + \frac{1}{2}MV_c^2$$

where I_c is the moment of inertia of the body about the axis of rotation. (*Hint*: Use the methods of Section 12-5, but recall that the velocity of the ith particle is now $(\mathbf{v}_i + \mathbf{V}_c)$, where $\mathbf{v}_i$ is the velocity of the ith particle due to the rotation of the body about its axis.)

32. Consider, in Figure 12-27, a spool of thread of mass M and radius b. If it rolls without slipping on a horizontal surface as it unwinds under the action of a horizontal force F_0, calculate:

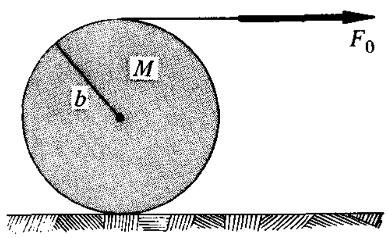


Figure 12-27

(a) The acceleration of the center of mass of the spool.
(b) The force that the surface exerts on the cylinder.

33. Consider, in Figure 12-28, a partially unwound spool of thread of total mass M and of radius b. If the moment of inertia of the spool and thread about its center of mass is I_c, the radius of the wound-up thread is c, and the thread is being unwound by a horizontal force F_0, find the acceleration of the center of mass assuming that the spool rolls without slipping. Express your answer in terms of F_0, M, b, c, and I_c.

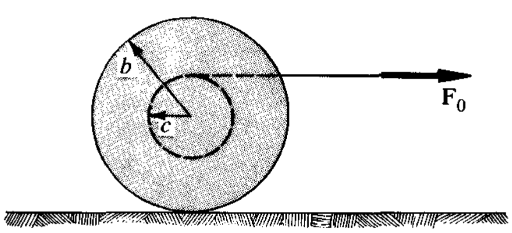


Figure 12-28

34. A homogeneous cylinder of radius b and mass M rolls, without slipping, down an inclined plane of angle β. Define x to be the distance of the center of mass along the plane and θ the angle through which the cylinder has turned.
   (a) Write down the implications of Theorems I and II for this case. Let N and R represent the normal and the tangential components of the force that the inclined plane exerts on the cylinder.
   (b) Determine the acceleration of the center of mass down the inclined plane in terms of the given parameters.
   (c) Calculate N and R.

35. For the physical situation described in Problem 34:
   (a) Define a zero of potential energy and write down the energy integral. Assume that the cylinder starts out at rest at the top of the inclined plane.
   (b) Calculate the velocity of the center of mass of the cylinder after it has traveled a distance l down the inclined plane.

*36. Consider two spheres of the same mass M and the same radius b, but suppose that one of these is hollow and the other is solid and homogeneous. Suppose that both are released from rest at the top of an inclined plane of angle β and that they both roll without slipping. Show that the ratio of the velocities of the two spheres when they reach the bottom of the plane is $(25/21)^{1/2}$. Which one arrives first?

*37. A very small, homogeneous sphere of radius b and mass M rolls, without slipping, down an inclined plane of angle β and onto a circular track of radius a (see Figure 12-29). Assume that the sphere starts from rest at the top.
   (a) Write down the conservation-

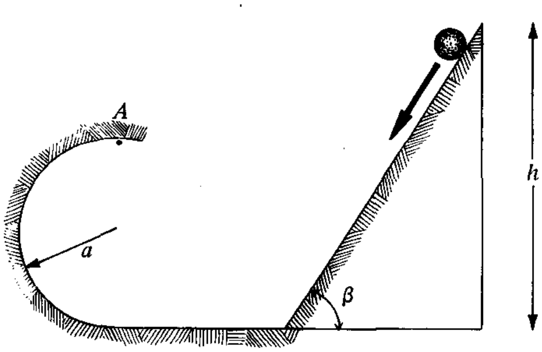


Figure 12-29

of-energy relation for times while the sphere is still on the inclined plane.

(b) What is the angular velocity of the sphere when it reaches the bottom of the inclined plane?

(c) Show that if $b \ll a$, then the sphere will reach point A, provided that the inequality

$$h \geq \frac{27a}{10}$$

is satisfied.

38. Imagine a star with a mass twice that of our sun and with a radius the same as the solar radius (Appendix B) and rotating about its axis with a period of 30 days. If this star were to collapse to form a *neutron star*, with a radius $R = 30$ km, what would its period of rotation become?

39. Consider again the situation in Figure 12-20. If the moment of inertia I_0 of the wheel about its axis is 1.2 kg-m^2, its angular velocity $\omega_0 = 20\pi$ rad/s, the moment of inertia I of the man, the platform, and the disk is 10 kg-m^2, calculate the final angular velocity ω of the platform for the situation shown in Figure 12-20b. Calculate also the change in the kinetic energy of the system. What is the source of this energy?

40. If, for the situation shown in Figure 12-19, the parameter values are $\omega_0 = 10$ rad/s, $I_0 = 15$ kg-m^2, and $I = 10$ kg-m^2, calculate **(a)** the final angular velocity ω of the system and **(b)** the kinetic energy lost or gained by this system.

41. A man of mass m stands at the center of a horizontal, circular turntable, of radius R_0, and moment of inertia I, and rotating on frictionless bearings at the angular velocity ω_0 about a vertical axis through its center. If the man suddenly walks radially outward until he is at the edge of the turntable, show that the final angular velocity ω of the system is

$$\omega = \frac{\omega_0}{1 + mR_0^2/I}$$

***42.** A man of mass m initially stands at the edge of a horizontal and round turntable of moment of inertia I_0 and of radius R_0. Assume that the platform is initially at rest but is mounted on frictionless bearings and that it is free to rotate about a vertical axis through its center. Suppose that the man starts walking around the edge of the platform at a velocity v_0 relative to it.

(a) Show that the angular velocity of the platform will be

$$\omega_0 = \frac{mR_0v_0}{I_0 + mR_0^2}$$

(b) What is the velocity v of the man relative to the ground?

(c) Through what angle will the turntable have rotated when the man returns to his initial position relative to ground?

13 Pressure in fluids—I

For the molecules of the body are indeed so numerous and their motion so rapid that we can perceive nothing more than average values.

LUDWIG E. BOLTZMANN (1844–1906)

. . . When it is not in our power to determine what is true we ought to follow what is most probable.

RENÉ DESCARTES (1596–1650)

13-1 General introduction

We turn now from our studies of mechanics and, with this chapter, begin a consideration of two types of macroscopic physical systems not previously encountered. These systems are *liquids* and *gases*. Unlike rigid bodies, whose dynamics can be described in their entirety by use of Theorems I and II, the dynamical behavior of liquids and gases is not generally susceptible to direct analysis by use of Newton's laws.

A collection of a very large number of molecules is said to constitute a *gas*, provided it is characterized by a density of the order of 10^{25} molecules/m^3 or less. Ordinary air is a typical example. It is a general property of a gas that it tends to expand and thereby to occupy all space available to it. A confined gas

will thus invariably fill up and assume the shape and the volume of its confining vessel.

A *liquid* is similar to a gas in that it also contains a very large number of molecules, but it differs in that its density is typically a thousand times higher. For example, there are approximately 3×10^{28} water molecules (H_2O) in a cubic meter of water, whereas there are only about 10^{25} of these molecules in a cubic meter of steam. A liquid may therefore be thought of as a very dense gas. Generally speaking, liquids also tend to assume the geometric shape of their container, but on the other hand they do not have the same tendency, as do gases, to occupy all available space. Rather, in a liquid there is a tendency for the constituent molecules to stick close together. Thus, there exist droplets of a liquid, but there are no droplets of a gas. Sometimes, we shall use the term *fluid*—that is, a substance that flows—when referring to either a liquid or a gas.

In general, it is not practical to describe in detail the dynamical behavior of a liquid or a gas by analyzing the Newtonian equations of motion for each particle. There is simply no hope of solving this system of equations, to say nothing of writing it down. Furthermore, how could the required initial positions and velocities for the 10^{25} molecules comprising the fluid ever be determined?

How then can we analyze the behavior of a system as complex as a liquid or a gas? The particular approach that is used in any given case depends, in general, on the physical and dynamical state of the system under consideration. In the following we shall be concerned principally with liquids and gases when they are in a certain quiescent state known as *thermal equilibrium*. For these cases, as we shall see, it is possible to analyze the system and to make certain quantitative predictions of its behavior. The term "thermal equilibrium" will be defined later. For the moment, we can think of it as that final state achieved by any system after it has been in contact with its environment for a long enough time so that processes involving energy transfer to and from the environment no longer take place. A cup of hot coffee, for example, comes into thermal equilibrium with the cool air in a room (its environment) by giving up energy to the air and thus cooling off. The discipline concerned with the study of macroscopic systems in thermal equilibrium is known as *thermodynamics*; it is the subject matter of this and the next four chapters.

According to the laws of mechanics, in order to describe the state of a gas consisting of $N\,(\sim 10^{25})$ molecules, it is necessary to specify $6N$ parameters: the initial position and velocity of each molecule, for example. By contrast, the description of the state of this system when it is in thermal equilibrium necessitates the specification of only two parameters: its volume and energy; or its volume and temperature; or its temperature and pressure. These various physical quantities, *volume, energy, pressure*, and *temperature*, in terms of which we describe the state of a system in thermal equilibrium, are known as *thermodynamic variables*. We shall begin our consideration of them in this and the next chapter by focusing on the thermodynamic variable of pressure.

13-2 Density of a dilute gas

One of the very simplest physical systems that can be analyzed by the use of the laws of thermodynamics is the *ideal* or *perfect* gas. Qualitatively speaking, an ideal gas is the limiting state approached by an ordinary gas as its density becomes very small. In order to give a more precise definition of the ideal gas, in this section we shall examine a gas at the microscopic level and thereby develop a criterion by means of which we can quantify this notion.

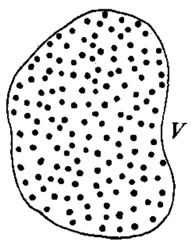


Figure 13-1

Consider, in Figure 13-1, a gas containing N molecules and confined to a closed container of volume V. We define the *particle density*, or simply the density, n of the gas by

$$n = \frac{N}{V} \tag{13-1}$$

so that n represents the number of molecules in a unit volume of a gas. A quantity related to the particle density is the *mass density* ρ, and this is defined by

$$\rho = mn = m\frac{N}{V} \tag{13-2}$$

where m is the mass of a gas molecule. In the next section we shall argue that for a gas in thermal equilibrium, if the dimensions of the container are not excessively large, say of the order of a meter or so, the density of the gas will be uniform throughout the container. Under conditions of thermal equilibrium, then, the number of particles per unit volume is the same in each part of the container, and a single number n serves to characterize the density of the gas as a whole.

Since the volume V of a container is easily measured, in order to obtain the density of a gas contained within it, it is necessary to know the number N of molecules in the gas. For this purpose, a knowledge of Avogadro's number plays an indispensable role. Let us recall from the discussion in Section 1-10 that Avogadro's number N_0 represents the number of molecules in a mole of a substance. It has the numerical value

$$N_0 = 6.02 \times 10^{23} \text{ molecules/mole} \tag{13-3}$$

Since a mole represents the number of grams of a substance numerically equal to its molecular weight, it follows that 18 grams of water (or of steam or of ice) or 2 grams of hydrogen or 4 grams of helium each consists of 6.02×10^{23} molecules of H_2O or H_2 or He, respectively. Therefore, if the total mass of a gas and its molecular weight are known, then the number of moles of the gas is also known, and by use of (13-3) we may then calculate N, the number of molecules in the gas.

Table 13-1 gives values for the densities and the mass densities of a variety of substances. In each case the particle density n is obtained by dividing the mass density ρ by the molecular mass m in accordance with (13-2). (For air, we have assumed an average molecular weight of 29.) Note that the densities of liquids and solids are typically 1000 times those of gases. The fact that the particle densities of the gaseous substances air, helium, and carbon dioxide all have the same value, 2.7×10^{25} molecules/m^3, is not a coincidence. We shall see later that at 0°C and at a pressure of 1 atmosphere (1 atm) all dilute gases have the same density of 2.7×10^{25} molecules/m^3.

Table 13-1 Densities of certain substances (at 0°C and at a pressure of 1 atmosphere)

Substance	*State*	*Density* n (molecules/m^3)	*Mass density* ρ (kg/m^3)
Air	Gas	2.7×10^{25}	1.3
Helium	Gas	2.7×10^{25}	0.18
Carbon dioxide	Gas	2.7×10^{25}	2.0
Mercury	Liquid	4.2×10^{28}	1.4×10^4
Water	Liquid	3.3×10^{28}	1.0×10^3
Ice	Solid	3.1×10^{28}	9.2×10^2
Aluminum	Solid	6.0×10^{28}	2.7×10^3
Iron	Solid	8.4×10^{28}	7.8×10^3

To define more precisely what is meant by an ideal gas, consider in Figure 13-2 a typical curve for the force F between two molecules separated by a distance r. This particular curve represents the force between two helium atoms as first determined theoretically by Slater and Kirkwood[1] and is

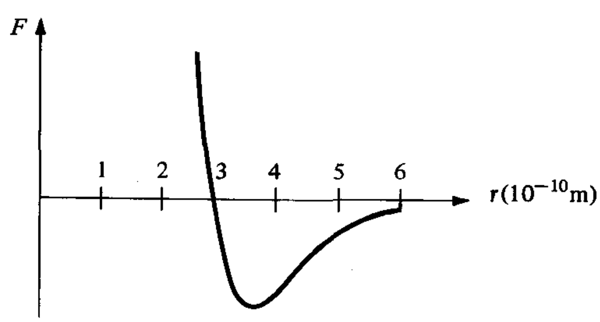


Figure 13-2

[1]J. C. Slater and J. G. Kirkwood, *Physical Review*, vol. 37, p. 682, 1931.

representative of the force between most atoms and many simple molecules. It is very similar to the force in Figure 8-15b, which is derived from the Lennard–Jones 6–12 potential in (8-19). For separation distances less than 3×10^{-10} meter, it follows from Figure 13-2 that the force between two atoms is very strong and repulsive, whereas for separation distances larger than this, the force is relatively weak and attractive. In particular, for distances greater than about 5×10^{-10} meter, the force between the atoms is negligible.

Let us define the *force range*, a, to be the smallest separation distance between two atoms for which they do not experience each other's forces. According to Figure 13-2, we have then

$$a \cong 5 \times 10^{-10}\ \text{m} \tag{13-4}$$

Thus when two atoms are separated by a distance greater than this force range a in (13-4), they are uncorrelated in that they exert no force on each other.

To define the ideal gas more precisely, consider the N molecules of a gas occupying a vessel of volume V. If each atom is thought of as a sphere of radius equal to the force range a, then the space actually occupied by the gas molecules is $(4\pi a^3/3)N$. The gas is said to be ideal if this occupied space is very small compared with the volume of the container. In other words, the gas is *ideal* if the inequality

$$\frac{4\pi}{3}a^3N \ll V$$

is satisfied. Making use of (13-1) and substituting the value for the force range a in (13-4), we conclude that for values of n satisfying the relation

$$n \ll 2 \times 10^{27}\ \text{atoms/m}^3 \tag{13-5}$$

on the average the atoms are outside of each other's force range, and under these circumstances the gas is ideal.

Reference to Table 13-1 shows that for the three gaseous substances there the density is 2.7×10^{25} molecules/m^3. Since this is very much less than 2×10^{27} molecules/m^3, to a fair approximation we may think of these gases as being ideal. Indeed, this is why all three have the same particle density. Note also from the table that the densities of liquids and solids do not satisfy (13-5).

Example 13-1 On a certain rocket flight, a measurement of the density of the atmosphere at an elevation of 60 km above the surface of the earth yielded the value 3.8×10^{-4} kg/m^3. Assuming an average molecular weight for air of 29, find the particle density n of the atmosphere at this altitude.

Solution Solving (13-2) for n and substituting the known values for ρ and m, we obtain

$$n = \frac{\rho}{m} = \frac{3.8 \times 10^{-4}\ \text{kg/m}^3}{29 \times 1.67 \times 10^{-27}\ \text{kg}} = 7.8 \times 10^{21}\ \text{molecules/m}^3$$

Since this is appreciably less than the value in (13-5), it follows that the air at this altitude is an ideal gas.

13-3 Microscopic view of an ideal gas

In this section we shall view an ideal gas at the microscopic level and present physical arguments to show that when in thermal equilibrium it has the following properties:

1. The magnitudes of the velocities of its constituent molecules are comparable to each other.
2. The particle density n is the same throughout the confining vessel.
3. The directions of travel of the molecules are more or less random throughout the gas.

To establish these characteristics of the ideal gas, consider in Figure 13-1 such a gas of N molecules in thermal equilibrium and confined to a container of volume V. Let us single out a typical molecule—that is, a molecule that has all the attributes of the "average" molecule in the gas—and focus attention on it as it moves about the container. Suppose, first, that it is traveling at some definite velocity $\mathbf{v}_0$ when it strikes one of the restraining walls of the container. One and only one of three things can happen to it during this collision. It can either lose energy, or it can gain energy, or it can have an elastic collision with the wall. If it were to lose (or gain) energy to the wall it follows that the gas as a whole would be losing (or gaining) energy from its surroundings, since by hypothesis the molecule is typical. But a gas in thermal equilibrium is one that, by definition, is not gaining or losing energy at the expense of its surroundings; so this is not possible. Therefore the molecule can only have an elastic collision when it strikes the wall. In actuality, any given molecule can lose (or gain) energy on colliding with the walls, but this must always be compensated for by a corresponding gain (or loss) of energy by another molecule so that, from the overall point of view, the gas maintains its state of thermal equilibrium. Figure 13-3 depicts a typical molecule with velocity $\mathbf{v}_0$ undergoing an elastic collision with a wall of the container. As shown, the magnitude of the velocity of the molecule, v_0, must be the same before and after the collision, as must be the angle θ which these two velocity vectors make with the normal to the wall.

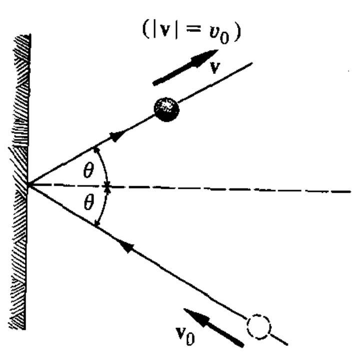


Figure 13-3

Let us continue to follow this typical molecule as it now reenters the gas. Since there are no external forces acting on it, it will continue to travel with its postcollision velocity $\mathbf{v}$ until it strikes either another wall or a second gas molecule. In this latter case, which arises if the separation distance between two molecules becomes less than the force range, we say that a *molecular collision* has taken place. In general, both the magnitude and the direction of the velocity of the molecule will change as a result of such a collision, and this then provides the mechanism by means of which the molecules exchange energy and momentum. Without molecular collisions it would not be possible for a gas to come to thermal equilibrium.

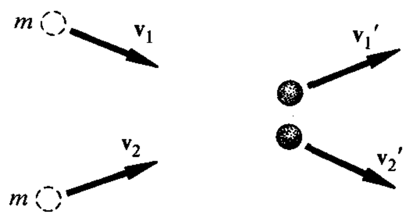


Figure 13-4

Figure 13-4 shows two colliding molecules, each of mass m and with initial velocities $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively. The corresponding velocities $\mathbf{v}_1'$ and $\mathbf{v}_2'$ after the collision are, as we know, not arbitrary but are related to the precollision values by the laws of momentum and energy conservation:

$$m\mathbf{v}_1 + m\mathbf{v}_2 = m\mathbf{v}_1' + m\mathbf{v}_2' \tag{13-6}$$

and

$$\frac{1}{2}m\mathbf{v}_1^2 + \frac{1}{2}m\mathbf{v}_2^2 = \frac{1}{2}m\mathbf{v}_1'^2 + \frac{1}{2}m\mathbf{v}_2'^2 \tag{13-7}$$

For given values for the initial velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ these conservation laws do not, in general, determine unique values for the final velocities $\mathbf{v}_1'$ and $\mathbf{v}_2'$. However, an analysis of collisions of this type shows that, in general, the faster-moving molecule has a tendency to lose energy to its slower partner. In other words, as a consequence of molecular collisions there is a tendency for the kinetic energies of the particles to become equalized. Imagine, for example, the extreme case where one of the colliding partners is originally at rest; after the collision, the kinetic energy of this molecule will, in general, not be zero. These arguments suggest the following property of a dilute gas in thermal equilibrium:

Property I

As a consequence of intermolecular collisions, there is a tendency for the molecules of the gas to acquire and to maintain velocities whose magnitudes are more or less comparable to one another.

Note that this does not mean that the molecules all tend to have precisely the same kinetic energy. Rather it means that it would be unusual to find a molecule with a velocity appreciably different in magnitude from that of a typical or average molecule. This average velocity about which the velocities of the molecules in the gas tend to aggregate is called the *thermal velocity*, or the *root-mean-square velocity*. It will be defined more precisely in the next section.

A second important property of a dilute gas confined to a container is that at thermal equilibrium its density n is uniform throughout. In macroscopic terms this means that on the average there are approximately the same number of molecules in equal volumes of the gas.

To see this, suppose that there were a region in the container in which the density of the gas were much higher than average. When an additional molecule enters this region, its chances of undergoing a collision here are greater than elsewhere in the container. Therefore, not only is it more likely that the original molecule will be scattered out of this denser region but its colliding partner will have a tendency to leave this region as well. This then is a mechanism that tends to bring about a depletion of molecules from regions of high density and an associated buildup of molecules in low-density ones. Thus we have made plausible the following property of a dilute gas:

Property II

As a result of intermolecular collisions, there is a tendency for molecules originally in regions of high density to go into regions of lower density. Thus the density of the gas tends to become uniform throughout the container.

Note that this does not mean that all regions of the container must invariably have the same density. Indeed, unusual situations in which the gas might, for example, collapse and occupy only half of the container can, in principle, occur. However, such inhomogeneities will not persist since molecular collisions will inevitably tend to reestablish the uniformity of the gas density everywhere in the container.

Making use of the fact that the density of a gas is uniform, we can now also make plausible the fact that the distribution of velocities of the molecules must be isotropic. Physically this means that, on the average, for every molecule entering a small volume element ΔV of the gas there must be a molecule with the same velocity leaving it. For consider a molecule of velocity $\mathbf{v}$ as it enters the given volume element ΔV and adds an amount of momentum $m\mathbf{v}$ to the gas inside. If there were not a compensatory leak of momentum out, there would be an accumulation of momentum, and thus the time rate of change of momentum of the gas inside ΔV would not be zero. This would lead to the manifestly incorrect result that the center of mass of the gas

inside ΔV accelerates. We conclude therefore that the existence of a preferred direction of travel in the gas is incompatible with the condition of thermal equilibrium. Hence follows:

Property III

To be consistent with the fact that there can be no accumulation of momentum in any volume element of the gas, the distribution of the velocities of the molecules must be isotropic.

Example 13-2 One mole of ^{4}He is confined to the left half of a rigid container of total volume 2.0×10^4 cm^3 by a rigid partition, as in Figure 13-5a. Suppose that the partition is suddenly removed.

(a) Describe qualitatively what happens.

(b) Find the initial and the final values for the density in the *occupied* regions of the container.

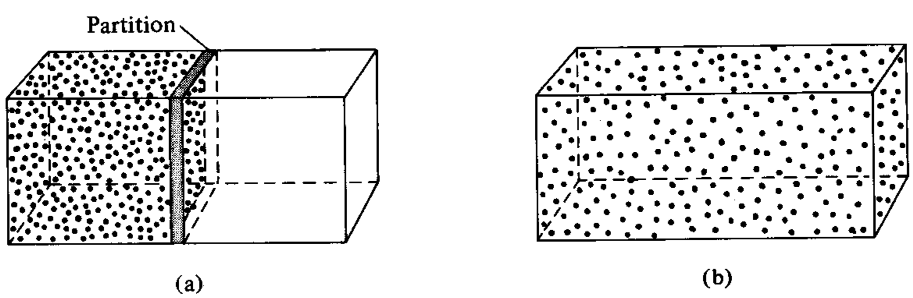


Figure 13-5

Solution

(a) Initially the gas is confined to a container of volume 10^4 cm^3 and is in thermal equilibrium. Immediately after the partition is removed, the gas is no longer in thermal equilibrium for it finds itself in a very inhomogeneous state, with a high density in the left half of the container and with zero density on the right. Because of Property III the molecules are moving isotropically in all directions, so immediately after removal of the partition some of them will stream into this vacuous region. Ultimately, as we saw above, because of molecular collisions the gas will become uniformly distributed throughout the entire container and come into thermal equilibrium (see Figure 13-5b).

(b) Since there are 6.0×10^{23} molecules in one mole, the initial gas density n_i is

$$n_i = \frac{N}{V} = \frac{6.0 \times 10^{23} \text{ atoms/mole}}{1.0 \times 10^{-2} \text{ m}^3} = 6.0 \times 10^{25} \text{ atoms/m}^3$$

where we have used the fact that there are 10^6 cm^3 in 1 cubic meter. Similarly, after thermal equilibrium is reestablished in the larger volume the final density n_f is

$$n_f = \frac{6.0 \times 10^{23} \text{ atoms/mole}}{2.0 \times 10^{-2} \text{ m}^3} = 3.0 \times 10^{25} \text{ atoms/m}^3$$

13-4 The thermal velocity

Consider a dilute gas containing N molecules and confined to a volume V. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ represent the instantaneous velocities of these molecules at some time t. We define the *root-mean-square velocity* or the *thermal velocity*, v_{th}, of the gas to be the positive square root of the average of the squares of these velocities; that is,

$$v_{th}^2 = \frac{1}{N}[\mathbf{v}_1^2 + \mathbf{v}_2^2 + \cdots + \mathbf{v}_N^2] \tag{13-8}$$

and thus v_{th} is a measure of the average velocity of the molecules in the gas. It follows from the arguments of Section 13-3 that as a result of molecular collisions the magnitudes of the velocities of the molecules of a gas in thermal equilibrium tend to acquire and to maintain a value near v_{th}.

In order for the thermal velocity, v_{th}, to be a meaningful gas parameter it is necessary that at thermal equilibrium it be independent of time. The fact that v_{th} has this property can be easily established in the following way. The total kinetic energy KE of the gas at any time t is

$$KE = \frac{1}{2} m [\mathbf{v}_1^2 + \mathbf{v}_2^2 + \cdots + \mathbf{v}_N^2]$$

and this may be expressed in terms of v_{th} by use of (13-8) as

$$KE = \frac{1}{2} m N v_{th}^2 \tag{13-9}$$

Assuming, then, that the gas is very dilute, so that on the average the potential energy between the molecules is negligible, it follows that the total energy[2] E of the gas is all kinetic and thus the kinetic energy must itself be a constant of the motion. Therefore, since in (13-9) m and N are fixed parameters, it follows that v_{th} is also constant in time.

In order to derive a second important property of the thermal velocity of a dilute gas of N molecules, let us set up a Cartesian coordinate system and let $v_{1x}, v_{2x}, \ldots, v_{Nx}$ be the components of the velocities of the molecules along the x-direction, and $v_{1y}, \ldots$ and $v_{1z}, \ldots$ be the corresponding components along the y- and z-directions, respectively. Define the average value $(v_x^2)_a$ by

$$(v_x^2)_a = \frac{1}{N}[v_{1x}^2 + v_{2x}^2 + \cdots + v_{Nx}^2] \tag{13-10}$$

and in a similar way the average values $(v_y^2)_a$ and $(v_z^2)_a$. Making use of the definition of the thermal velocity in (13-8), we have

$$v_{th}^2 = (v_x^2)_a + (v_y^2)_a + (v_z^2)_a \tag{13-11}$$

[2]As will be discussed more fully in Section 16-7, although the conclusion is correct, this argument is applicable only to a monatomic gas.

Moreover, according to Property III the distribution of velocities must be isotropic for gas in thermal equilibrium. Therefore, since there can be no preferred direction for the velocities, it follows that

$$(v_x^2)_a = (v_y^2)_a = (v_z^2)_a \tag{13-12}$$

Finally, then, combining (13-11) with (13-12) we find that

$$(v_x^2)_a = (v_y^2)_a = (v_z^2)_a = \frac{1}{3}\, v_{\text{th}}^2 \tag{13-13}$$

a result that we shall need later.

Example 13-3 Consider again the mole of ^{4}He confined to the left half of a rigid container of total volume $2.0 \times 10^4\ \text{cm}^3$ in *Figure* 13-5a and suppose that $v_{\text{th}} = 2.0 \times 10^5$ cm/s. Assume that the collisions with the walls of the container are elastic.

(a) Calculate the initial energy of the gas.

(b) Calculate the thermal velocity of the gas after the partition is removed.

Solution

(a) Since the gas is dilute, its energy is entirely kinetic. Substituting the given parameter values into (13-9), we obtain

$$\begin{aligned} KE &= \frac{1}{2}\, mNv_{\text{th}}^2 = \frac{1}{2} \times 6.67 \times 10^{-27}\ \text{kg} \times 6.0 \times 10^{23} \times (2.0 \times 10^3\ \text{m/s})^2 \\ &= 8.0 \times 10^3\ \text{J} \end{aligned}$$

(b) Since, by hypothesis, the collisions with the walls are elastic and since no energy loss is associated with the collisions of the molecules among themselves, the total energy of the gas does not change as a result of removing the partition. Therefore, since the gas is dilute, its kinetic energy is conserved and thus, according to (13-9), v_{th} has the same value of 2×10^5 cm/s after the expansion that it had before.

The process shown in Figure 13-5, in which a gas is allowed to expand into a vacuum, is known as the *free expansion* of the gas. In the free expansion of a dilute gas the energy, and therefore also the thermal velocity, is not changed. However, experiments carried out at higher densities, say $n \approx 10^{27}$ molecules/m^3 show that the thermal velocity of a gas actually *decreases* slightly during a free expansion. Can you suggest a reason for this?

13-5 Pressure in a gas

Consider, in Figure 13-6, a rigid cubical box containing a dilute gas and let us focus attention on a molecule that is in the process of making an elastic collision with a wall S. For simplicity assume that the incident velocity $\mathbf{v}_0$ is perpendicular to S. After the collision the velocity of the molecule will be $-\mathbf{v}_0$, and the change in its momentum $\Delta\mathbf{p}$ as a result of this collision is

$$\begin{aligned} \Delta\mathbf{p} &= m\mathbf{v}_f - m\mathbf{v}_i \\ &= -2m\mathbf{v}_0 \end{aligned}$$

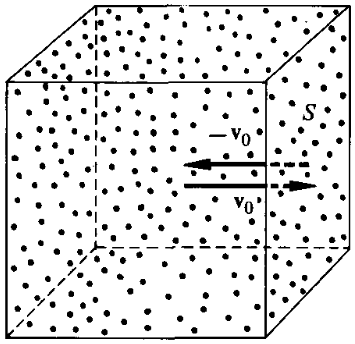


Figure 13-6

where $\mathbf{v}_i = \mathbf{v}_0$ is the velocity of the molecule before the collision and $\mathbf{v}_f = -\mathbf{v}_0$ is its velocity afterward. If we let τ represent the very short time interval that the molecule is in contact with the wall, then the impulsive force $\mathbf{f}_p$ experienced by the molecule during the collision is

$$\mathbf{f}_p = \frac{\Delta \mathbf{p}}{\tau} = -\frac{2m\mathbf{v}_0}{\tau}$$

Note that this force $\mathbf{f}_p$ is directed perpendicularly inward toward the interior of the container. According to Newton's third law, the molecule must exert on the surface a force equal and opposite to $\mathbf{f}_p$. Thus, as a result of this collision, the wall S of the container experiences a force

$$\mathbf{f} = -\mathbf{f}_p = \frac{2m\mathbf{v}_0}{\tau} \tag{13-14}$$

and this corresponds to a perpendicular outward push on S.

Viewing the gas as a whole now, we note that there will be a large number of molecules constantly colliding with the walls and, in accordance with (13-14), these produce a net outward push on each wall of the container. This total force is not localized as is the force $\mathbf{f}$ due to a single molecule. Instead, each surface of the container will experience a more or less uniform outward force, or in other words, from a macroscopic point of view each portion of the surface S experiences a steady and uniform force produced by the continuous bombardment of a very large number of molecules. This force is always directed perpendicularly outward from the surface of the container.

For a given particle density n, this total force $\mathbf{F}$ on a surface of the container must be proportional to the area A of the wall under consideration. For according to Properties II and III, the number of particles striking a wall or a part of a wall per unit area per unit time—and thus exerting a force on it—must be the same for each interior surface of the container. With this in mind, we define the *pressure*, P, of the gas to be the force per unit area exerted on any wall by the gas; that is,

$$P = \frac{F}{A} \tag{13-15}$$

where F is the net outward force exerted on a wall or a part of a wall of area A. Since the density and the distribution of velocities of the molecules of the gas are the same everywhere in the container, it follows from our picture of this force being produced by molecular collisions with the wall, that the pressure P must be the same on *all* walls of the container. Thus, even though the walls of the container may be of various shapes and sizes, so that the force which the gas exerts on different walls is different, the pressure P is the same on each wall.

More generally, suppose that a container of arbitrary shape, such as in Figure 13-7, contains a dilute gas. If P is the gas pressure, then the force $\mathbf{F}_1$ exerted on any element of surface ΔA_1 by the gas is given by the product $P\ \Delta A_1$ and is directed outward and perpendicular to the area element itself. A similar result follows for any other element such as ΔA_2.

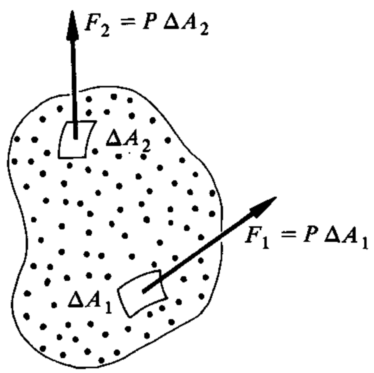


Figure 13-7

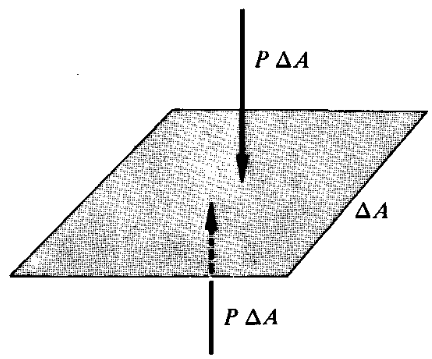


Figure 13-8

For many purposes it is convenient to think of a gas pressure as existing in the interior of a gas as well. Consider, in Figure 13-8, a small surface element ΔA in the interior of a vessel which contains a gas at a pressure P. The density and velocity distribution of the molecules in the neighborhood of the surface will be the same as they are near the walls of the container. Accordingly, the gas molecules will strike and bounce from ΔA and therefore exert a force on it in the same way they do at the surface of the container. However, the gas exerts *no* net force on this area element ΔA in the interior of the gas. For the force $P\ \Delta A$ on one side will be balanced by an equal and opposite force of strength $P\ \Delta A$ from the other. This differs in a vital way from the situation at the walls, where the gas molecules strike only that side of the surface facing the gas. Thus even though the molecules of a gas exert no force on any surface in its interior, the gas exerts a pressure there nevertheless.

Making use of our physical picture of pressure, it is easy to establish that the pressure P of a gas in thermal equilibrium must have the same value at all interior points as it has on the surface. For if it did not, there would have to be an inhomogeneity in either the density or the velocity distribution of the gas, and this is not possible for a gas in thermal equilibrium.

Figure 13-9 depicts a simple way of measuring the pressure of a gas. A gas is

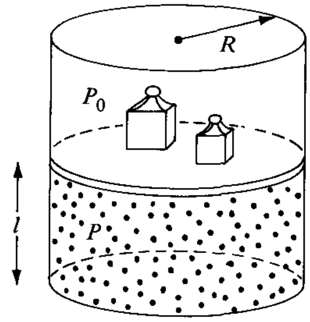


Figure 13-9

confined to a volume $\pi l R^2$ in a cylinder of radius R by means of a piston whose total weight W is adjustable by placing various auxiliary weights on top of it. If P is the pressure of the confined gas and P_0 is the atmospheric pressure of the air on top of the piston, then, because the piston is in mechanical equilibrium,

$$P\pi R^2 = W + P_0 \pi R^2$$

since the upward push $P\pi R^2$ due to pressure of the confined gas is balanced by the downward pressure of the atmosphere and the weight W of the piston. Solving for $(P - P_0)$, we obtain

$$P - P_0 = \frac{W}{\pi R^2}$$

Since both W and R are independently measurable, this provides a direct way for measuring the pressure P of the confined gas relative to the pressure P_0 of the atmosphere. This pressure difference $(P - P_0)$ is known as the *gauge pressure.*

The unit of pressure is force per unit area. Thus in SI units, this is the newton per square meter (N/m^2). The pound per square foot (lb/ft^2) is the unit pressure in the British engineering system. Another unit of pressure is the *atmosphere* (atm). This represents the pressure of our atmosphere at sea level and has the value $1.01 \times 10^5\ N/m^2$.

Example 13-4 Consider a gas confined to a container by means of a piston. Suppose a perpendicular force $F_0 = 10$ newtons must be exerted on the piston to keep the gas from expanding. If the area of the piston is $20\ cm^2$, what is the gas pressure?

Solution The pressure of the gas is the force per unit area acting on the piston. Therefore

$$P = \frac{F_0}{A} = \frac{10\ N}{20\ cm^2}$$
$$= 0.5\ N/cm^2 = 5.0 \times 10^3\ N/m^2$$

Note that this value for P is not the total pressure of the gas but only the gauge pressure. The total pressure is obtained by adding to P the atmospheric pressure produced by the collision of air molecules in the surrounding atmosphere with the "external" face of the piston.

Example 13-5 Consider the free expansion of the gas in Figure 13-5. If P_0 is the initial pressure of the gas when it occupies the volume 10^4 cm^3, what is its pressure P when it occupies the full volume of 2.0×10^4 cm^3?

Solution After the free expansion, the density of the gas is halved but the thermal speed v_{th} is not altered. Consequently, on the average, the walls of the container will be bombarded with half as many molecules per unit time after the expansion as they were before. It follows from the arguments presented in this section that the gas pressure must, of necessity, be halved. That is, since the pressure on the walls is determined by the number of molecules per unit time striking them, it follows that if the density is halved, only half as many are available near the walls at any given time. Therefore

$$P = \frac{1}{2} P_0$$

Note that if v_{th} had changed during the expansion—so that it would not have been a free expansion—then although the density would have been halved, the pressure would *not.* For even though, on the average, half as many particles would be bombarding the walls of the container after the expansion they would be doing so with a different thermal velocity; and this would also affect the pressure.

This example is easily generalized to arbitrary volume changes. Using the same arguments as above we conclude that during a free expansion the pressure of a dilute gas varies inversely with volume. If P_0 and V_0 are the initial values for the pressure and volume of a gas prior to a free expansion, and P and V are the corresponding values afterward, then

$$PV = P_0 V_0 \qquad \text{(free expansion)}$$

This result is known as *Boyle's law.* Figure 13-10 shows a plot of P versus V for a free expansion.

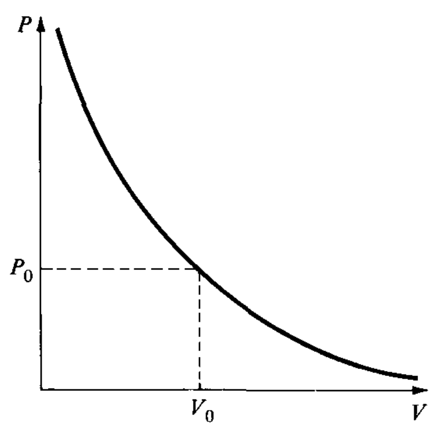


Figure 13-10

13-6 Pressure variation with thermal velocity

Based on the physical interpretation of pressure as given in Section 13-5, we expect that the pressure of an ideal gas will increase with both its density n

and its thermal velocity v_{th}. For the greater the number of molecules that strike the walls of the container and the greater the velocity with which they strike, the larger is the gas pressure. Detailed studies show that, for a dilute gas, the pressure P is directly proportional to the product of the density n and the square of the thermal velocity v_{th}^2. We shall now give a heuristic derivation of this striking and important relation and in the process obtain the correct coefficient of proportionality.

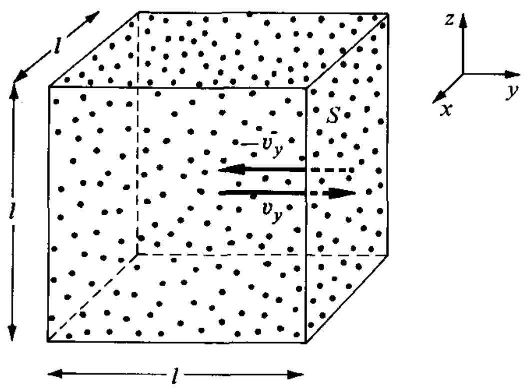


Figure 13-11

Consider a dilute gas of N molecules in thermal equilibrium and confined to a cubical container of side l, as in Figure 13-11. Let us focus attention on a single molecule, which in terms of the coordinate system shown in the figure is instantaneously moving along the positive y-axis with a velocity $\mathbf{j}v_y$. (The essentials of the following argument are not altered if this molecule also has components of velocity along the other two axes.) As shown in the figure, after the molecule strikes the surface S it is reflected back into the gas with a velocity $-\mathbf{j}v_y$. A time interval l/v_y later, the molecule will strike the surface opposite to S, be reflected back into the gas with a velocity $\mathbf{j}v_y$, directed to the right this time, and finally after a total time interval of length $2l/v_y$ it will return to S. There it will be reflected back into the gas a third time. In other words, if there are no collisions with the other molecules, the given molecule will strike S once during each time interval of length $2l/v_y$. At each such collision with S it suffers a momentum change $\Delta\mathbf{p} = -2\mathbf{j}mv_y$ and thus undergoes a change in momentum per unit time, $\Delta\mathbf{p}/\Delta t$, given by

$$\frac{\Delta\mathbf{p}}{\Delta t} = \frac{-2\mathbf{j}mv_y}{2l/v_y} = -\mathbf{j}\frac{mv_y^2}{l}$$

According to Newton's second and third laws, the force $\mathbf{f}$ that this molecule exerts on S is

$$\mathbf{f} = \mathbf{j}\frac{mv_y^2}{l}$$

and the total force $\mathbf{F}$ that all the gas molecules exert on S is obtained by adding

together the forces produced by each of them. Thus

$$\mathbf{F} = \mathbf{j}\frac{m}{l}[v_{1y}^2 + v_{2y}^2 + \cdots + v_{Ny}^2]$$
$$= \mathbf{j}\frac{m}{l}N(v_y^2)_a$$

where the second equality follows by use of (13-10). Substituting for $(v_y^2)_a$ in terms of v_{th}^2 in accordance with (13-13), we find that the total force $\mathbf{F}$ on S is

$$\mathbf{F} = \frac{1}{3}\mathbf{j}\frac{m}{l}Nv_{th}^2 \tag{13-16}$$

As anticipated above, this force $\mathbf{F}$ in (13-16) is perpendicular to S and is directed outward from the container. Since the area of the surface S is l^2, by use of the definition in (13-15) the gas pressure P is

$$P = \frac{F}{l^2} = \frac{1}{3}\frac{m}{l^3}Nv_{th}^2$$
$$= \frac{1}{3}\frac{N}{V}mv_{th}^2 \tag{13-17}$$

where in the third equality we have used the fact that the volume V of the gas is l^3. According to (13-1) and (13-2), equivalent forms for P are

$$P = \frac{1}{3}nmv_{th}^2$$
$$= \frac{1}{3}\rho v_{th}^2 \tag{13-18}$$

Also, since the total energy E of a dilute gas is the same as its kinetic energy in (13-9) another way to express (13-18) is

$$PV = \frac{2}{3}E \tag{13-19}$$

Example 13-6 A sample of air at sea level has a mass density of 1.3 kg/m^3 (see Table 13-1) and a measurement of its pressure yields the value 1.0×10^5 N/m^2. Calculate the thermal velocity v_{th} of air molecules under these circumstances.

Solution Solving (13-18) for v_{th}, we find that

$$v_{th} = [3P/\rho]^{1/2}$$

and substituting into this the given values for P and ρ, we obtain

$$v_{th} = \left[\frac{3 \times 1.0 \times 10^5\ \mathrm{N/m^2}}{1.3\ \mathrm{kg/m^3}}\right]^{1/2} = 480\ \mathrm{m/s}$$

Generally speaking, this value is typical for the thermal velocity of a gas at room temperature.

Example 13-7 Suppose that in Figure 13-6 a gas molecule is traveling at a speed of 500 m/s back and forth between the two parallel surfaces of the container separated by a distance of 0.4 meter. How many times will this molecule collide with the side S in 1 second? Assume that it does not collide with any other molecule.

Solution For the given velocity, the total distance traveled by the molecule in 1 second is 500 meters. Since the distance it must travel between each collision with S is 2×0.4 meter = 0.8 meter, it follows that it makes 500 meter/0.8 meter = 625 collisions with this wall per second. This means that during the time interval it takes to carry out a macroscopic measurement on the gas—this is typically of the order of 1 second—the molecules of the gas go back and forth across the container 1250 times!

13-7 Pressure and density variations in the atmosphere

In the above discussion of the pressure and density of a dilute gas we have assumed that gravitational effects are negligible, so that at thermal equilibrium both P and n are uniform throughout the container. In this section we examine this problem in more detail and show that this assumption is justified, provided that the container is not too large.

Consider, in Figure 13-12a, a very thin, horizontal slab of our atmosphere of area A and thickness Δz, and at a height z above the surface of the earth. Let $\rho(z)$ and $P(z)$ represent the mass density and the pressure, respectively, of the air at this elevation. Since the mass in this volume element is $\rho(z)A\ \Delta z$, the downward pull of gravity on the air in it is $\rho(z)A\ \Delta z g$. The only other forces acting on this slab are those produced by the pressure due to the surrounding air molecules. As shown in Figure 13-12b, the air molecules above the slab produce on it a downward force of strength $P(z+\Delta z)A$, while those beneath the slab produce an upward force $P(z)A$. Since the air in this volume element can be taken to be in equilibrium, it follows that

$$P(z)A - P(z+\Delta z)A - \rho(z)A\ \Delta z\, g = 0 \tag{13-20}$$

which, for the special case $g = 0$, reduces to $P(z+\Delta z) = P(z)$, as it must if it is to be consistent with our previous discussions.

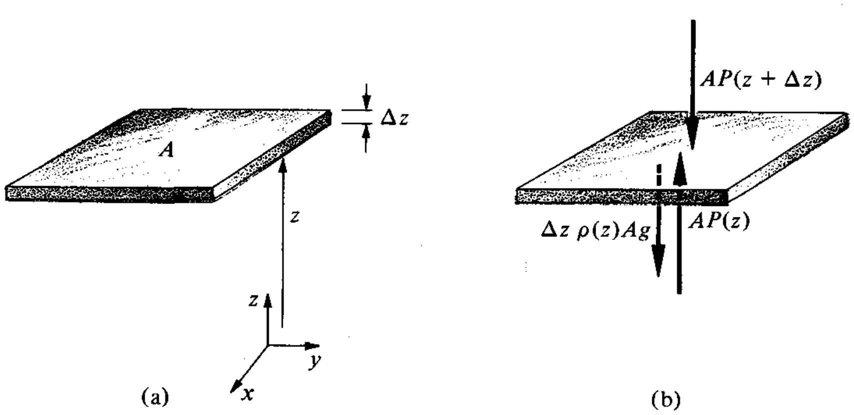


Figure 13-12

To analyze (13-20), let us now specialize to the case of an infinitesimally thin slab by taking the limit $\Delta z \to 0$. According to the definition of the derivative, for all sufficiently small Δz we have

$$\frac{P(z+\Delta z)-P(z)}{\Delta z} \cong \frac{dP}{dz}$$

with the equality more exact the smaller is Δz. Substituting this form for $P(z+\Delta z)$ into (13-20), we find that the terms $AP(z)$ cancel and, if we cancel out the common factor $A\,\Delta z$ from the remaining terms, there results

$$\frac{dP}{dz} = -\rho(z)\,g \tag{13-21}$$

To obtain an explicit formula for the pressure P as a function of elevation, (13-21) must be supplemented by a relation of some type involving the density $\rho(z)$. Let us assume that our atmosphere may be described as an ideal gas, so that P and ρ are related as in (13-18) at each elevation z. Solving for ρ and substituting into (13-21), we find that

$$\frac{dP}{dz} = -\frac{P}{z_0} \tag{13-22}$$

where z_0 is a parameter known as the *scale height* and is defined by

$$z_0 = \frac{v_{\text{th}}^2}{3g} \tag{13-23}$$

Physically, z_0 represents the distance over which the pressure and the density of the atmosphere vary appreciably. Using the values $v_{\text{th}} = 500$ m/s and $g = 9.8$ m/s^2, we obtain $z_0 \cong 8$ km. In other words, for changes in elevation that are very small compared to a kilometer there is no appreciable variation of the atmospheric pressure or the density.

Because of heating effects associated mainly with the sun, the thermal velocity of the air in our atmosphere is generally different at different elevations. For the moment, however, let us neglect this variation and assume that v_{th} is constant throughout the atmosphere. Then the scale height z_0, as defined in (13-23), is independent of elevation and it is easy to verify that for this case the solution of (13-22) is

$$P(z) = P_0 e^{-z/z_0} \tag{13-24}$$

with P_0 the atmospheric pressure at sea level (at $z = 0$). Experimentally, we find that at 0°C, $P_0 = 1.01 \times 10^5$ N/m^2. The corresponding formula for the mass density $\rho(z)$ may be obtained by substituting (13-24) into (13-18). The result is

$$\rho(z) = \rho_0 e^{-z/z_0} \tag{13-25}$$

where ρ_0 is the density of the atmosphere at sea level and is related to P_0 in accordance with (13-18).

Figure 13-13 shows a plot of the variation of the pressure in the atmosphere

as a function elevation z, as predicted by (13-24). In the problems it will be established that the scale height z_0 may be interpreted as that height above sea level at which the pressure and the density of the atmosphere have dropped to $1/e \cong 0.368$ of their values at sea level. As shown in the figure, at an elevation of one scale height, that is, for $z = z_0$, the pressure of the atmosphere is P_0/e.

For thermal velocities of the order of 500 m/s, as we saw above, the scale height z_0 is enormous—of the order of 8 km. This means that for dilute gases in containers with dimensions in the order of, say, 1 meter, we may to a very high degree of precision neglect the effects of gravity. This, then, justifies our previous neglect of gravitational effects and our continuing to neglect these effects in the following.

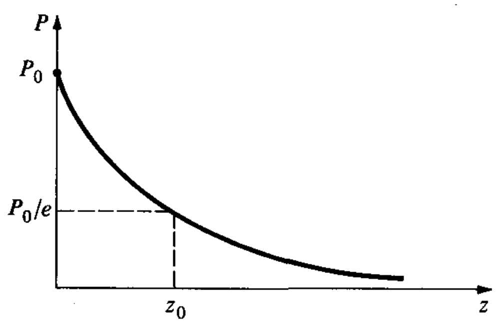


Figure 13-13

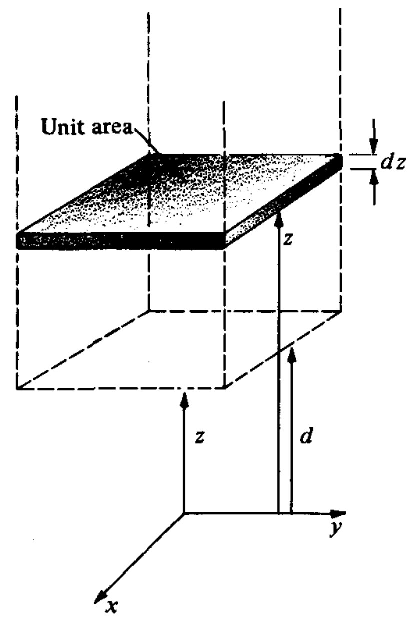


Figure 13-14

Example 13-8 Show that the atmospheric pressure at a vertical distance d above sea level is numerically equal to the weight of all of the air in the vertical column of unit area above this elevation.

Solution Let us first calculate the total amount of gas in a vertical column of unit area above the elevation d. Consider, in Figure 13-14, a thin slab of air of unit area and of thickness dz at an elevation z ($> d$) above sea level. The mass dM of this slab is

$$dM = \rho(z)\,dz = \rho_0 e^{-z/z_0}\,dz$$

where we have used (13-25). By integration, we find that the total mass M of the air in this column is

$$M = \int_d^\infty \rho_0 e^{-z/z_0}\,dz = \rho_0[-z_0 e^{-z/z_0}]\Big|_d^\infty$$

$$= \rho_0 z_0 e^{-d/z_0} = \frac{P_0}{g}\,e^{-d/z_0}$$

where the last equality follows by use of (13-23) and (13-18). The total weight W of this column of air is thus

$$W = Mg = P_0 e^{-d/z_0}$$

and the desired result then follows by comparison with (13-24).

13-8 Work carried out on a gas

Consider, in Figure 13-15a, a volume Ax of a gas under a pressure P confined to a rectangular box with a movable piston of area A. Provided that an external agent exerts a force PA on this piston, the gas will remain in thermal equilibrium and no motion of the piston will occur. However, if the agent were to exert a force greater than PA, the force on the piston would be unbalanced and the piston would move in such a direction as to compress the gas. Correspondingly, if the force due to the external agent were to become less than PA, then the net force on the piston would be unbalanced in the opposite direction and the gas would expand. In the former case, where the gas is compressed, we say that *positive work* has been carried out on the gas by the agent, and in the other case, where the gas expands, we say that *negative work* has been carried out on the gas. Our choice of sign in this connection is arbitrary; indeed, some authors use the opposite convention. In all of the following, positive work will invariably mean that work is being carried out *on* the gas.

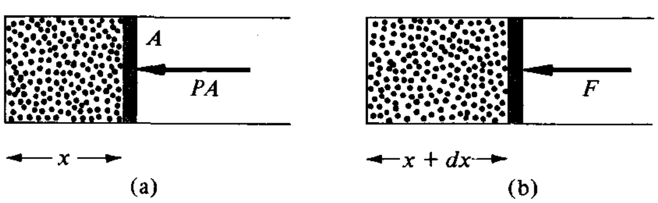


Figure 13-15

The actual work carried out by an agent in compressing a gas depends in general on the strength of the force that he exerts and the distance through which he displaces the piston. Consider as a special case the situation shown in Figure 13-15b, where the force F that the external agent exerts on the piston is slightly less than the force due to the gas pressure PA. As the volume of the gas changes from its original value, Ax, to its final one, $A(x + dx)$, the work dW carried out by the agent on the gas is $-F\,dx$, where the negative sign reflects the fact that the force and the displacement of the piston are in opposite directions. Since F differs only very slightly from the force PA exerted on the piston by the gas, the work dW carried out on the gas is

$$dW = -PA\,dx$$

Equivalently, since in a displacement of the piston by the distance dx the

volume V of the gas is changed by $dV = A\,dx$, we may write

$$dW = -P\,dV \tag{13-26}$$

The minus sign in this formula is required by our sign convention. If $dV < 0$, then, according to (13-26), $dW > 0$ and work is carried out *on* the gas. Conversely, if $dV > 0$, then $dW < 0$ and work is carried out on the external agent by the gas. Even though (13-26) was derived for a rectangular container, this formula is independent of geometry and is very generally valid.

To illustrate the usage of (13-26) let us calculate the work W carried out on an ideal gas as it is compressed from an initial volume V_i to a final one V_f. To be able to make use of (13-26) we must think of the gas as being compressed in a series of small steps so that at each of these steps the gas is in thermal equilibrium. If we were to squeeze down on the gas very suddenly and with a very large force, for example, then density and pressure inhomogeneities would be generated in the gas and it would not be in thermal equilibrium during and immediately after the compression. Under this circumstance the concept of pressure would simply not be applicable. Thus we must imagine the gas as being compressed very slowly and in very small steps, so that after each step the gas can return to its quiescent state of thermal equilibrium. In this way the pressure P of the gas remains a valid concept, and the work W carried out on the gas in compressing it from a volume V_i to V_f is

$$W = -\int_{V_i}^{V_f} P\,dV \tag{13-27}$$

In order to evaluate (13-27) we need to know how the gas pressure P varies as a function of the volume. In general, even assuming that the gas remains in thermal equilibrium, this variation depends on how the compression is carried out. To be specific, let us assume that the gas is ideal and that the compression is carried out in such a way that the thermal velocity v_{th} is not changed during the compression. Substituting (13-17) into (13-27), we find that

$$\begin{aligned} W &= -\frac{1}{3}\,mNv_{th}^2 \int_{V_i}^{V_f} \frac{dV}{V} \\ &= -\frac{1}{3}\,mNv_{th}^2 \ln V \Big|_{V_i}^{V_f} \\ &= \frac{1}{3}\,mNv_{th}^2 \ln \frac{V_i}{V_f} \end{aligned} \tag{13-28}$$

As we might expect on physical grounds, if $V_f < V_i$, this work W is positive, whereas if $V_i < V_f$, W is negative. This is consistent with our sign convention in accordance with which the agent carries out positive work when the gas is compressed.

We emphasize that in deriving (13-28) we have assumed that v_{th} is constant during the compression. This is one of the simplest assumptions that could be

made. The fact that, in general, the thermal velocity need not be constant during a compression is illustrated in Example 13-11.

In Figure 13-16 we plot P as a function of volume V for fixed N and v_{th} in accordance with (13-17). This graph is known as an *isotherm.*[3] At the initial and final values for the gas volumes, V_i and V_f, respectively, we have constructed vertical lines to the isotherm. According to a basic property of an integral, the work W carried out on the gas is numerically equal to the shaded area under the curve between these two vertical lines. More generally, whether the gas is dilute or not, the work carried out in compressing it is numerically equal to the area under a suitable pressure–volume curve. If the gas expands, then in accordance with (13-27), W is the negative of the shaded area.

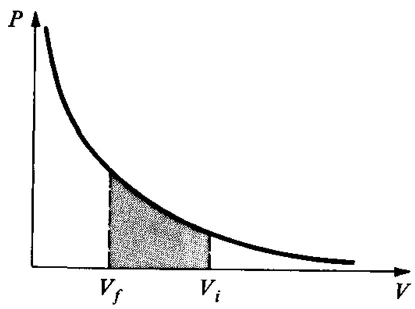


Figure 13-16

Example 13-9 Suppose a gas is compressed from an initial volume V_0 to a final volume $V_0/2$ in such a way that at each stage PV = constant. If P_0 is the initial value of the gas pressure, how much work is carried out on the gas?

Solution On the basis of the given information, at each stage of the compression we have

$$PV = P_0 V_0$$

Substituting this form into (13-27), we find that

$$W = -\int_{V_0}^{V_0/2} P\,dV = -P_0 V_0 \int_{V_0}^{V_0/2} \frac{dV}{V}$$

$$= -P_0 V_0 \ln \frac{V_0/2}{V_0} = P_0 V_0 \ln 2$$

where the final equality follows from the fact $\ln 1/x = -\ln x$.

Example 13-10 An *isobaric compression* is one in which a gas in compressed at constant pressure. Calculate the work carried out on a gas at atmospheric pressure in compressing it isobarically from an initial value $V_i = 2.0 \times 10^3\ \text{cm}^3$ to a final value $V_f = 1.5 \times 10^3\ \text{cm}^3$.

[3]As will be defined in Chapter 15, the term "isotherm" refers to the fact that the temperature of the gas is the same at all points on this curve.

Solution For this case, where the pressure is constant, P may be taken out from under the integral sign in (13-27). Hence

$$W = -P\int_{V_i}^{V_f} dV = -P(V_f - V_i)$$
$$= (-1.0 \times 10^5\ \mathrm{N/m^2}) \times (1.5 \times 10^{-3}\ \mathrm{m^3} - 2.0 \times 10^{-3}\ \mathrm{m^3})$$
$$= +50\ \mathrm{J}$$

where we have used the value $1.0 \times 10^5\ \mathrm{N/m^2}$ for the atmospheric pressure.

In a P-V diagram, such as the one shown in Figure 13-17, an isobaric process is represented by a horizontal line. The work carried out on the gas in such a process is given by the rectangularly shaded area.

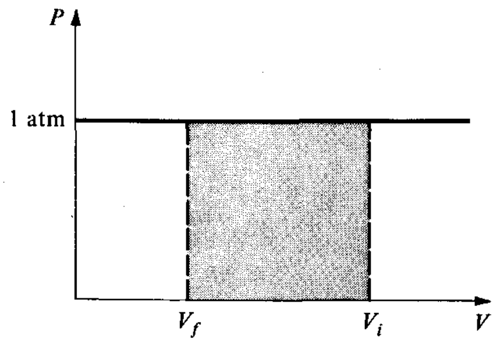


Figure 13-17

Example 13-11 Suppose that a molecule of mass m is traveling at a velocity $v_0(\cong v_{th})$ when it strikes at right angles a piston that is moving toward the molecule at a velocity u. Calculate the change in the energy ΔE, of the molecule, assuming that the resulting collision would have been elastic if the piston were not moving.

Solution Figure 13-18a shows the situation just prior to the collision. To calculate the velocity with which the molecule rebounds from the piston after the collision, let us view the collision from the viewpoint of a coordinate system at rest with respect to the piston. As shown in Figure 13-18b, from this point of view the molecule approaches the piston with a velocity $(v_0 + u)$. Since the collision is elastic, afterward it will have a velocity $(v_0 + u)$, directed toward the left, away from the piston. Finally, in Figure 13-18d we see the postcollision situation from the viewpoint of the original observer, with respect to whom the molecule has the final velocity $(v_0 + 2u)$.

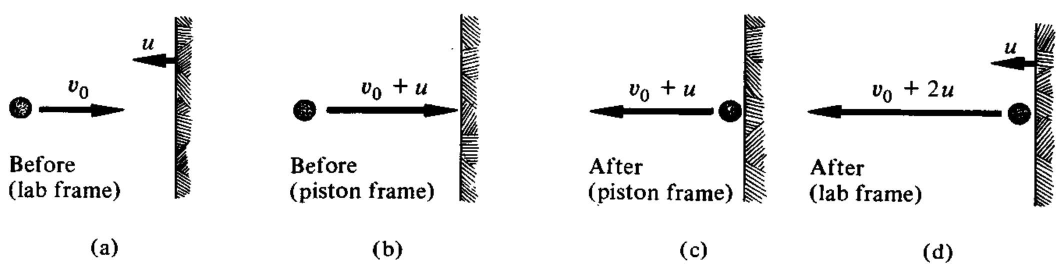


Figure 13-18

The energy of the molecule just prior to the collision is $\frac{1}{2}mv_0^2$, and afterward its energy is $\frac{1}{2}m(v_0+2u)^2$. Accordingly, the energy change ΔE is

$$\Delta E = \frac{1}{2}m(v_0+2u)^2 - \frac{1}{2}mv_0^2$$
$$= 2mv_0u + 2mu^2$$

For the normal situation, where $v_0 \cong v_{th} \cong 10^3$ m/s and $u \ll v_{th}$, this may be written as

$$\Delta E \cong 2mv_0u$$

If u is positive, as in Figure 13-18a, so that the piston moves toward the molecule before the collision, then ΔE is positive and the energy of the gas is increased. Correspondingly, if $u < 0$, so that before the collision the piston is moving away from the molecule, then ΔE is negative.

We can conclude from this example that if a piston moves into a dilute gas, the total energy of the gas will *increase* by virtue of the fact that the molecules colliding with the piston will in general have their energy increased. Correspondingly, when the piston moves out, the energy of the gas is decreased. By use of (13-27) this result may be interpreted by saying that the thermal velocity of the gas will increase or decrease, respectively, depending on whether the piston is moving into or out of the gas.

13-9 Summary of important formulas

The thermal velocity v_{th} of a gas consisting of N molecules is defined by

$$v_{th}^2 = \frac{1}{N}[\mathbf{v}_1^2 + \mathbf{v}_2^2 + \cdots + \mathbf{v}_N^2] \tag{13-8}$$

where $\mathbf{v}_i$ is the velocity of the ith molecule. A gas is said to be an *ideal gas* provided that it is sufficiently dilute that on the average the molecules are not subject to one another's force range to any appreciable extent. The kinetic energy of an ideal gas is

$$KE = \frac{1}{2}mNv_{th}^2 \tag{13-9}$$

and, since the potential energies between the molecules can be neglected, in this case this is the same as the total energy E of the gas.

For an ideal gas the pressure P and the thermal velocity v_{th} are related by

$$P = \frac{1}{3}mnv_{th}^2$$
$$= \frac{1}{3}\rho v_{th}^2 \tag{13-18}$$

where $\rho = mn$ is the mass density of the gas.

The infinitesimal work dW carried out on a gas under a pressure P by an agent that "slowly" changes its volume by an amount dV is

$$dW = -P\,dV \tag{13-26}$$

where the sign is such that, for a compression, positive work is carried out on the gas. For a finite change, the corresponding formula is

$$W = -\int_{V_i}^{V_f} P\,dV \tag{13-27}$$

provided that the work is carried out in such a way that at each stage the gas is in thermal equilibrium.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: (a) ideal gas; (b) force range; (c) thermal equilibrium; (d) free expansion; (e) isotherm; and (f) gauge pressure.
2. Give three examples of systems that are in thermal equilibrium with their environment. Give three examples of systems that are *not* in thermal equilibrium with their surroundings. Is a pot of water boiling on a stove in thermal equilibrium?
3. Explain in physical terms why it is that the air inside a balloon is uniformly distributed throughout its interior. What modifications, if any, are required in your argument to account for the fact that the density of the air in our atmosphere decreases with elevation?
4. A *plasma* is a gas containing charged particles for which the interparticle forces vary inversely with the square of the distance. Does our criterion defining an ideal gas apply to this case? Explain.
5. Consider again the system in Figure 13-5 for producing the free expansion of a gas. Suppose this time, however, that the density n of the gas in the container is of the order of 10^{27} molecules/m^3. Will v_{th} for the gas increase or decrease as a result of the free expansion? Explain your answer in terms of the intermolecular force curve in Figure 13-2.
6. Explain why the density of an ideal gas confined to a container need not be uniform if its thermal velocity v_{th} vanishes. (*Hint*: What velocity will each of the molecules have in this case?)
7. For a gas of identical molecules in thermal equilibrium, why must the vector sum of the velocities of all the molecules vanish? Is this inconsistent with the fact that the thermal velocity of such a gas has a nonzero value?
8. In the collision of two identical molecules in a gas, is it always necessary for the slower moving one to gain energy? If your answer is no, explain why there is nevertheless a tendency for the kinetic energies of the molecules to become equalized.
9. Explain in physical terms the mechanism by means of which a gas produces pressure. Why must the pressure produced by a gas in thermal equilibrium be uniform throughout the gas?
10. Explain why the pressure P of a gas must always be positive. What physical meaning would a negative pressure have?
11. An ideal gas of N molecules is confined to a volume V. Suppose, as a result of the application of a very large force to a piston, that the volume of the gas is suddenly halved. After thermal equilibrium is reestablished, which of the quantities N, V, n, v_{th}, and P have decreased? Which have increased? Which will remain the same? In each case justify your answer in physical terms.
12. Repeat all parts of Question 11, but

suppose this time that the gas is suddenly expanded to twice its volume.

13. In terms of the intermolecular force field, explain why it is that when a molecule of a dense gas strikes the surface of a container it will, in general, exert a smaller force on it than it would if it were a molecule in a dilute gas. In other words, explain in physical terms why there is a small pressure drop associated with a rise in the density of a gas.
14. In light of the fact that a moving piston will change the energy of a gas, explain how it is possible to compress a dilute gas without changing its thermal velocity.
15. Consider the horizontal slab of air in Figure 13-14. Explain in physical terms why you would expect the air molecules above the horizontal slab to exert a smaller pressure on it than do those beneath it.
16. Describe an experiment by means of which you could measure the thermal velocity of the air molecules around you.
17. Explain in physical terms how it is possible to carry out the isobaric expansion of a dilute gas in Example 13-10. Does the thermal velocity increase or decrease in such an expansion?
18. Two dilute gases occupy the same volume V. If P_1 and P_2 would have been their respective pressures if each had been the sole occupant of the volume, explain in physical terms under what circumstances you would expect the total pressure of the gas to be $(P_1 + P_2)$.
19. Is it possible to have the "free compression" of a dilute gas? Why is such a phenomenon never observed?
20. Why does it make no sense to use the definition for work carried on a gas in (13-27) unless the gas is in thermal equilibrium throughout the process?

PROBLEMS

1. What is the mass density ρ of a sample of gaseous carbon dioxide which has a particle density $n = 2.3 \times 10^{25}$ molecules/m^3? Can we consider this sample to be an ideal gas?
2. A mass of 3 grams of ^{4}He occupies a volume of 2.0×10^4 cm^3. **(a)** What is the mass density of the gas? **(b)** Calculate the particle density n of this gas. Can we consider the sample to be an ideal gas?
3. Using the data in Table 13-1, calculate, for the stated pressure and temperature, the volume occupied by 1 mole of **(a)** air (average molecular weight 29); **(b)** carbon dioxide; and **(c)** helium. **(d)** Compare your answers to (a), (b), and (c). Would you expect a similar result to hold for any dilute gas?
4. A liter is a unit of volume defined to be 10^3 cm^3. It is often stated that under the conditions of pressure and temperature in Table 13-1, 1 mole of a dilute gas occupies 22.4 liters. Using this fact, calculate the mass densities of **(a)** nitrogen (molecular weight 28); and **(b)** neon (atomic weight 20). Compare your results with those found in a handbook.
5. The lowest density of a substance that we know of occurs in interstellar space, where it is believed that there is, on the average, 1 hydrogen atom per cm^3. Calculate the values for n and ρ for this interstellar gas.
6. At the present time it is believed that the highest density that occurs in nature is in the nucleus of an atom. Assuming that a nucleus consisting of A neutrons and protons has a

radius r which is given (in meters) by

$$r = 1.1 \times 10^{-15} A^{1/3}$$

calculate n and ρ for nuclear matter. Compare this with the corresponding value for water.

7. Consider 1 cubic meter of air at the conditions of pressure and temperature in Table 13-1. Estimate the average distance between these air molecules by assuming that they are arranged in a regular way in a cubical volume 1 meter on a side.
8. Consider a system of six identical particles, each of mass 10 grams and traveling along the x-axis of a certain coordinate system at the respective velocities ± 2 m/s, ± 3 m/s, and ± 4 m/s.
   - **(a)** What is the average velocity of these particles?
   - **(b)** What is the thermal velocity v_{th} of these particles?
   - **(c)** Calculate the total kinetic energy.
9. Suppose that the thermal velocity of 1 mole of a dilute gas of hydrogen molecules is 520 m/s. **(a)** What is the kinetic energy E of this gas? **(b)** If the gas occupies a volume of 0.03 cubic meter, what is the gas pressure?
10. Calculate the thermal velocity v_{th} and the energy of 3 grams of ^4He confined to a volume of 2.0×10^4 cm^3 at atmospheric pressure.
11. One mole of a dilute gas of N_2 occupies a volume of 25 liters.
    - **(a)** If the gas pressure is 1 atm, what is the thermal velocity of the molecules?
    - **(b)** If the volume of the gas is decreased to 10 liters, find the new pressure of the gas, assuming that the thermal speed of the molecules is doubled in this process.
    - **(c)** What is the increase in the kinetic energy of the gas as a result of the compression in (b)?

*12. Consider a dilute gas of "hard-sphere" molecules, each of radius a.
   - **(a)** Find the volume swept out in a time interval τ by a molecule that has a velocity v_{th}.
   - **(b)** We define τ to be the *collision time*, provided that a molecule of the gas (other than the original one) is found in the volume calculated in (a). Show that the collision time τ is

     $$\tau = \frac{1}{n\pi a^2 v_{th}}$$

     where n is the particle density of the gas.
   - **(c)** The *mean free path* λ is defined to be the average distance traveled by a molecule at the velocity v_{th} before it suffers a collision. Show that λ has the value

     $$\lambda = \frac{1}{n\pi a^2}$$

13. Consider a gas of hard-sphere molecules, each of radius 2.0×10^{-10} meter and traveling at a thermal velocity of 4.0×10^4 cm/s. Making use of the results of Problem 12, calculate the mean free path λ and the collision time τ if:
    - **(a)** $n = 10^{26}$ molecules/m^3
    - **(b)** $n = 10^{25}$ molecules/m^3
    - **(c)** $n = 10^{18}$ molecules/m^3.

    In which of these three cases can a given molecule make, on the average, one traversal back and forth in a container with dimensions of the order of 1 meter without suffering a molecular collision? Can you suggest any reason why the collision time τ should be of the order of the time required for the gas to come to thermal equilibrium?
14. Consider, in Figure 13-11, a molecule traveling back and forth

horizontally at a thermal speed $v_{th} = 600$ m/s. Assume that gravity acts vertically downward, that is, along the negative z-axis, $l = 50$ cm, and that the collisions with the walls are elastic.

(a) How long does it take the molecule to go from one side of the container to the other?
(b) How far does the molecule drop under the action of gravity during this time interval?
(c) What is the velocity of the molecule after it has descended a vertical distance of 10 cm under the action of gravity?
(d) What must be the collision time τ so that the given molecule collides with another one before it falls a vertical distance of 0.02 cm under the action of gravity?

15. Consider in Figure 13-19 a container of total volume V partitioned into two equal volumes. Suppose that the left-hand side contains N_1 molecules, each of mass m, and the right-hand side contains N_2 identical molecules. Assume that the gases are dilute and characterized by the same thermal velocity v_{th}.

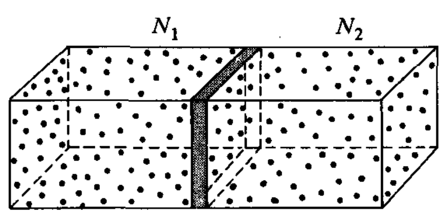


Figure 13-19

(a) What is the density and the pressure of each of the gases in terms of the given parameters?
(b) If the partition is suddenly removed, what are the final values for the particle density, the pressure, and the kinetic energy after thermal equilibrium is reestablished?

16. Consider again the physical situation depicted in Figure 13-19, but suppose this time that there are an equal number of molecules on both sides of the partition, that is $N_1 = N_2$, but suppose that the molecules on the two sides of the partition have different thermal velocities, $v_{th}^{(1)}$ and $v_{th}^{(2)}$, respectively.

(a) What are the initial energies E_1 and E_2 of the two gases in terms of the given parameters?
(b) If the partition is removed, why is the energy E of the gas after thermal equilibrium is reestablished the same as the sum of the energies calculated in (a)?
(c) By use of your result to (b) calculate the thermal velocity v_{th} of the gas after equilibrium is reestablished.

17. Consider the physical situation shown in Figure 13-19, with the following parameter values: $V = 5 \times 10^4\ \text{cm}^3$, $N_1 = 9 \times 10^{23}$, $N_2 = 3 \times 10^{23}$, and $m = 3.3 \times 10^{-26}$ kg. The thermal velocity of the molecules in the left-hand compartment is initially 500 m/s, and in the right it is 300 m/s.

(a) What are the initial values of the pressures in the two compartments?
(b) What is the kinetic energy of each of the gases initially?
(c) If the partition is removed, what is the thermal velocity and the pressure of the mixture after thermal equilibrium has been reestablished?

18. Suppose that the volume of the container in Figure 13-19 is 25 liters, and that the left-hand compartment has 0.5 mole of molecular oxygen, whereas the right-hand compartment has 0.7 mole of H_2. Assume that the pressures of both gases before the partition is removed is 1 atm.

(a) What are the initial values for the thermal velocity of each of the gases?
(b) Calculate the thermal velocity of the mixture after the partition is removed.
(c) Calculate the final pressure of the gas.

19. By use of (13-23) and (13-24), show that the scale height z_0 represents the height at which the pressure of our atmosphere has fallen to $1/e$ of its value at sea level.

20. If the scale height z_0 of our atmosphere is 8 km:
(a) What thermal velocity is associated with the air molecules in our atmosphere?
(b) Calculate the height below which one half of the mass of our atmosphere is found.
(c) Above what height would we find 10 percent of the mass of the atmosphere?

21. Assuming that the scale height of the atmosphere is 8 km and that at sea level the atmospheric pressure is 1.0×10^5 N/m^2 and the density is 1.3 kg/m^3, calculate the pressure and the density of the atmosphere at the following locations: (a) Mexico City (elevation 2.3 km); (b) the top of Pike's Peak (elevation 4.3 km); and (c) the top of Mount McKinley (elevation 6.7 km). Neglect in each case any possible variations due to temperature differences.

22. Suppose a gas is compressed along the solid straight-line path AB in the P-V diagram in Figure 13-20.
(a) In terms of the parameters P_A, P_B, V_A, and V_B, how much work is carried out on the gas in this process?
(b) How much work would be carried out on the gas if it were taken from point B to point A along the solid-line path BA?
(c) Express your answer to (a) in terms of an area in the P-V diagram in the figure.

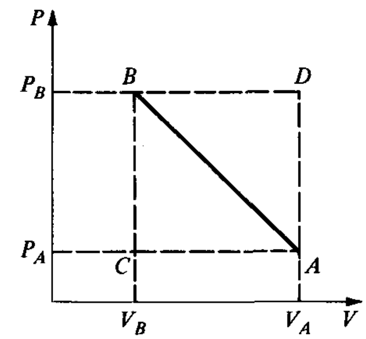


Figure 13-20

23. Consider again the gas undergoing the process in Figure 13-20.
(a) How much work is carried out on the gas if it is taken at constant volume from point A to point D?
(b) How much work is carried out on the gas if it is compressed isobarically from point D to point B?
(c) What is the total work carried out on the gas if it is taken from point A to point B by first taking it isobarically from A to point C, and then at constant volume V_B to point B?
(d) Interpret the above results in geometrical terms.

24. Consider the process depicted in Figure 13-20.
(a) How much work is carried out on the gas if it is forced around the rectangle $ADBCA$? What is the geometrical significance of your answer?
(b) Show that the work carried out in taking the gas around the closed triangle $ABCA$ is

$$W = \frac{1}{2}(P_B - P_A)(V_A - V_B)$$

(c) What would have been the work in (b) if the triangle were traversed in the opposite direc-

tion, namely along the path $ACBA$?

25. Generalize your result of Problem 24 by showing that if a gas is caused to go counterclockwise around an arbitrary closed cycle, as in Figure 13-21, the work carried out on the gas is numerically equal to the area enclosed by the path (the shaded area in the figure).

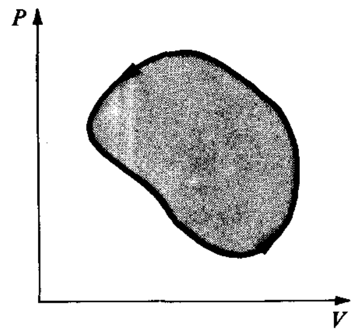


Figure 13-21

26. Consider one mole of ^{4}He, characterized by the pressure and volume associated with point A in Figure 13-22.

(**a**) What is the thermal velocity and energy of the gas at this point?

(**b**) If the gas is forced to go from point A to point B along the semicircular arc in the figure, calculate the amount of work carried out on the gas in this process.

(**c**) What is the energy and the thermal velocity of the gas at point B?

(**d**) Compare the energy gained by the gas with the work carried out on it. In physical terms, account

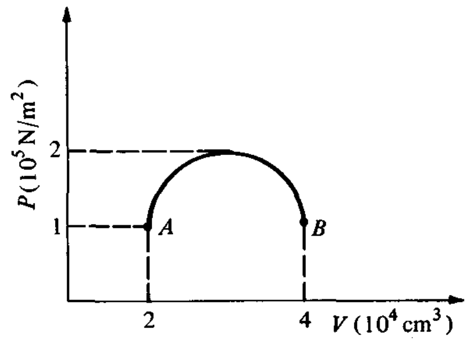


Figure 13-22

for any differences between these two quantities.

27. One mole of gaseous nitrogen is at a pressure of 1 atmosphere and occupies a volume of 22.4 liters. Calculate the work carried out on this gas if it is compressed isothermally (that is, in a way so that the product PV is constant throughout the process) to a final volume of (**a**) 20 liters and (**b**) 11.2 liters.

28. The potential energy curve $V(r)$ between two ^{4}He atoms is given, according to Slater and Kirkwood, by

$$\frac{1}{\epsilon}V(r) = 5.67 \times 10^6 \times \exp\left\{-21.5\frac{r}{\sigma}\right\} - 1.08\left(\frac{\sigma}{r}\right)^6$$

where $\sigma = 4.6 \times 10^{-10}$ meter and $\epsilon = 1.3 \times 10^{-23}$ joule.

(**a**) Make a plot of $V(r)$ as a function of the separation distance r. (**b**) Calculate the force $-dV/dr$ and determine the minimum of the force curve in Figure 13-2.

14 Pressure in fluids—II

There is nothing softer and weaker than water and yet there is nothing better for attacking hard and strong things.

LAO TZÉ (The Tao)

14-1 Introduction

We have previously observed that liquids and gases are similar in many respects, and they differ mainly in that liquids are generally several orders of magnitude more dense than are gases. Under the conditions in Table 13-1, for example, 1 cubic meter of air contains approximately 2.7×10^{25} molecules, whereas a corresponding volume of liquid water consists of 3.3×10^{28} H_2O molecules. However, this difference in density between liquids and gases has the important consequence that the macroscopic properties of these two classes of substances differ in a variety of important ways.

We begin our study of liquids in the next section by describing, in physical terms, the concept of pressure in a liquid and contrasting it with the related one of pressure in a gas. Next, we turn to a more quantitative formulation of this concept and apply it to a number of physical phenomena associated with liquids at rest. It will be apparent, as we go along, that many of these results apply equally well to gases, and the main purpose for confining the discussion to liquids is therefore only one of convenience.

In the concluding sections of this chapter we examine the concept of pressure in the more general framework of fluids in motion.

14-2 Liquid pressure-qualitative aspects

According to Figure 13-2 the force between any two atoms of a fluid is characterized by a force range $a \approx 5 \times 10^{-10}$ meter. If we take as typical the value $n = 3 \times 10^{28}$ molecules/m^3 for the particle density of a liquid, then a brief calculation shows that the density of a liquid may be characterized by

$$na^3 \approx 1 \qquad \text{(liquid)} \tag{14-1}$$

It is interesting to contrast this relation with (13-5); that is,

$$na^3 \ll 1 \qquad \text{(ideal gas)}$$

which characterizes the ideal gas. In physical terms, then, whereas for a dilute gas the volume occupied by the molecules, $4\pi Na^3/3$, is very small compared to the confining volume V, for a liquid the corresponding two volumes are comparable to one another. This means that for a liquid the intermolecular spacing is of the order of the force range, so that on the average each molecule experiences the attractive forces due to its nearest neighbors. By contrast, in a dilute gas the molecules are generally oblivious of each other's presence except during those rare occasions when they undergo a collision. And it is owing mainly to the greater effectiveness of the intermolecular force field in a liquid that its properties differ so markedly from those of a dilute gas.

Consider in Figure 14-1 a liquid, say water, which has been poured into a container. Unlike a gas, which expands to occupy all available space, the liquid will maintain its fixed density by filling the container from the bottom up and assume the shape of that portion of the vessel that it occupies. Because of the greater effectiveness of the attractive part of the force between molecules, this behavior of a liquid is just what we should expect on physical grounds. Just as for a gas, in general, the molecules of a liquid undergo motions and some of them may even acquire velocities comparable to those in a dilute gas. However, such fast-moving molecules will experience the attractive molecular forces due to their neighbors and thus will be inhibited from leaving the surface of the liquid. In other words, since the separation

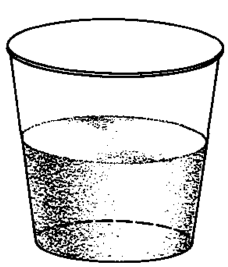


Figure 14-1

distances between the molecules of a liquid are on the average comparable to the intermolecular force range, the molecules of a liquid tend to stick together. Occasionally, a molecule will have a sufficiently large velocity to overcome the attractive forces of the other molecules and will be moving in just the right way to escape from the surface of the liquid. This process is known as *evaporation.* These escaping or evaporating molecules will in turn collide with the surrounding air molecules and some of them will eventually fall back into the liquid. Finally, when just as many molecules return to liquid per unit time as are evaporated from it, thermal equilibrium is established.

Based on this discussion we might expect that, despite the higher density, the mechanism underlying pressure in a liquid is very similar to that in a gas. This is indeed the case. Just as for a gas, we find that under ordinary circumstances the molecules of a liquid are constantly in a state of motion, so that when they collide with a wall of the confining vessel, they exert a force on it. *Hence a liquid exerts pressure by precisely the same mechanism that a gas does.* The basic distinction is that a liquid molecule which is in the process of striking the wall of the confining vessel will experience the attractive force from the main body of the liquid and will thus slow down to some extent. Hence, the pressure exerted by a liquid will, in general, be somewhat less than that exerted by a gas under similar circumstances.

There is one important way that the pressure variation in a liquid differs from that in a gas. This is due to the fact that pressure in a liquid is very much influenced by the force of gravity. Recall in this connection the discussion in Chapter 13, where we found that the pressure of a gas confined to a vessel with dimensions of the order of a meter is uniform throughout. By contrast, as we shall see quantitatively in the next section, the pressure in a liquid is very much dependent on gravity and in general varies throughout its volume.

14-3 Pressure in a liquid

Figure 14-2a shows a liquid confined to a vessel that is open at the top. Let P_0 represent the pressure of the atmosphere on the upper surface of the liquid,

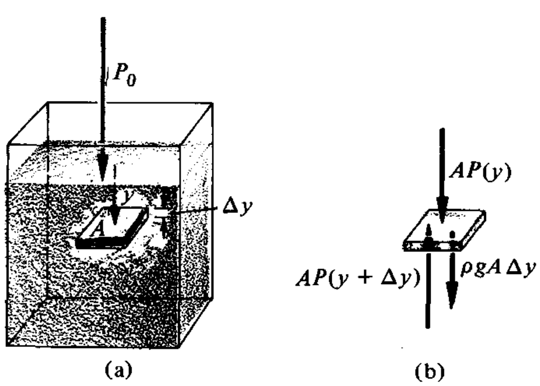


Figure 14-2

and let us consider an infinitesimal horizontal slab of the liquid of area A and thickness Δy at a vertical distance y *below* the surface. Figure 14-2b shows the three forces acting on this slab of fluid. In terms of the pressure $P(y)$ at the depth y and the fluid density $\rho(y)$ at this depth, these forces are (1) a downward force $AP(y)$ acting on the upper face of the slab; (2) an upward force $AP(y+\Delta y)$ due to the liquid at the lower surface of the slab; and (3) the downward gravitational pull $\rho g A\,\Delta y$. Since the slab of fluid is not accelerating, the vector sum of these three external forces acting on it must vanish. Hence

$$\rho g A\,\Delta y + AP(y) - AP(y+\Delta y) = 0 \tag{14-2}$$

Following the same procedure as used to derive (13-21), we find that in the limit as $\Delta y \to 0$, (14-2) becomes

$$\frac{dP}{dy} = \rho g \tag{14-3}$$

which is basically the same as (13-21) and differs only in sign by virtue of our choice of direction for the y-axis in Figure 14-2. This relation in (14-3) is basic to understanding the behavior of liquids at rest.

To obtain an explicit formula for the pressure variation in a liquid, it is necessary to supplement (14-3) by a second relation involving the pressure and the density. Unlike the analogous situation for a dilute gas—where it was assumed that the pressure and the density were proportional to each other in accordance with (13-18)—for a liquid this assumption is entirely unwarranted. Let us therefore restrict ourselves from now on to those physical situations for which the variation of the density of the liquid with pressure may be neglected. Any liquid that has this property is said to be *incompressible* and experiment shows that many liquids—ordinary water, for example—may be considered to be incompressible to a high degree of accuracy. Assuming then that ρ in (14-3) is a constant, independent of the pressure and therefore also of y, it follows that the derivative dP/dy is the constant ρg. Integrating, we find that

$$P(y) = \rho g y + P_0 \tag{14-4}$$

where P_0 is an integration constant and represents the value of P at the surface of the liquid at $y=0$. As shown in Figure 14-2a, P_0 also represents the atmospheric pressure. Why?

An equivalent way of expressing (14-4) may be obtained by reference to Figure 14-3. Let P_1 be the pressure at a point 1 in a liquid and P_2 the corresponding pressure at a second point 2, at a vertical displacement h above it. Applying (14-4) at both of these points and subtracting, we find that

$$P_1 - P_2 = \rho g h \tag{14-5}$$

or, in other words, the difference in pressure between any two points is proportional to their vertical separation. Note that, as shown in the figure, it is not necessary for these two points to lie along the same vertical line. The

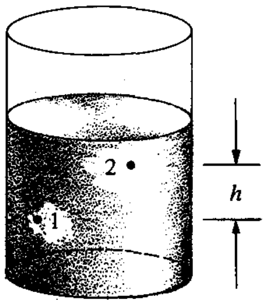


Figure 14-3

relation in (14-5) is applicable under all circumstances and (14-4) may be thought of as the special case for which the point 2 lies at the surface of the liquid.

An important feature of (14-4) is that the pressure varies only with vertical depth below the surface of the liquid and, in particular, is independent of the geometry of the containing vessel. Consider for example, in Figure 14-4, three vessels resting on a fixed horizontal surface, each one filled with the same fluid to a height h. Because of the fact that the pressure in a liquid depends only on the vertical distance below its surface, it follows that the pressure is the same at each of the four points A_1, A_2, A_3, A_4, all of which are at the same vertical distance below the horizontal surfaces of the liquids.

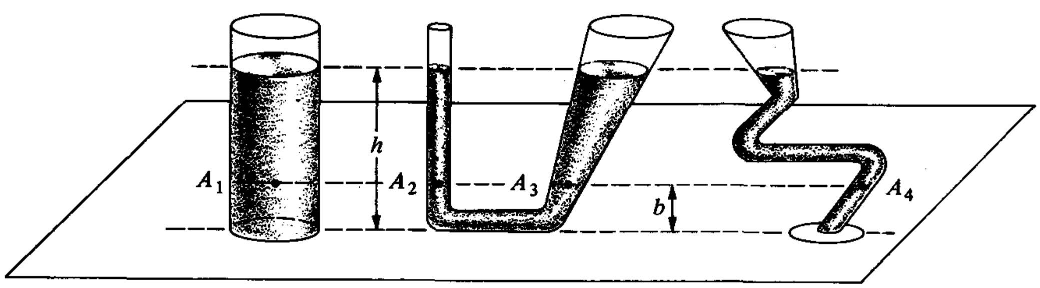


Figure 14-4

Another important consequence of the linear variation with depth of the pressure in a liquid in (14-4) is a relation first established experimentally by Blaise Pascal (1623–1662) and known as *Pascal's principle.* This principle states:

If the pressure at some point in an incompressible fluid is changed by some amount, then the pressure is changed by the same amount everywhere in the fluid.

The proof of this principle follows directly from (14-5). For since the pressure difference between any two points in a liquid depends only on the difference in their elevation, it follows that if the pressure at one point P_1 is

altered by some amount, then the pressure P_2 at any other point must be altered by precisely the same amount. Otherwise their difference would not have the constant value ρgh, as it must according to (14-5).

Example 14-1 Show that the pressure at depth d below the surface of a liquid is numerically equal to the weight of the column of the liquid (plus that of the atmosphere) of unit area above this depth.

Solution Consider the situation as shown in Figure 14-5. The weight of the column of liquid of unit cross-sectional area and of height d is ρgd, where ρ is the constant density of the liquid. In Example 13-8, we established that the atmospheric pressure at any fixed elevation, call it P_0, is equal to the weight of all the air in a column of unit area above this elevation. Thus the total weight W of the atmosphere and of the liquid contained in a column of unit area above the point at a distance d below the surface of the liquid is

$$W = P_0 + \rho gd$$

Finally, comparing this with (14-4) and setting $y = d$, we obtain the desired result.

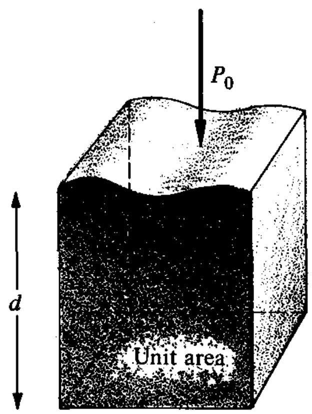


Figure 14-5

Example 14-2 Calculate the pressure at a depth of 50 meters below the surface of the ocean.

Solution For this case the parameter values are $\rho = 10^3\ \text{kg/m}^3$, $P_0 = 1.0 \times 10^5\ \text{N/m}^2$, and $y = 50$ meters. Substituting into (14-4) we find that

$$\begin{aligned} P &= P_0 + \rho gh \\ &= 1.0 \times 10^5\ \text{N/m}^2 + (10^3\ \text{kg/m}^3) \times (9.8\ \text{m/s}^2) \times 50\ \text{m} \\ &= 5.9 \times 10^5\ \text{N/m}^2 \end{aligned}$$

Thus the pressure at this depth is approximately six times that at the surface.

Example 14-3 Find the angle θ that the surface of a liquid in a container makes with the horizontal if the entire system is forced to accelerate along the horizontal direction with an acceleration $\mathbf{a}_0$.

Solution Figure 14-6a shows the physical system when it is not accelerating. Here, the surface of the liquid is horizontal; in other words, it is at right angles to the direction of the force of gravity on the liquid.

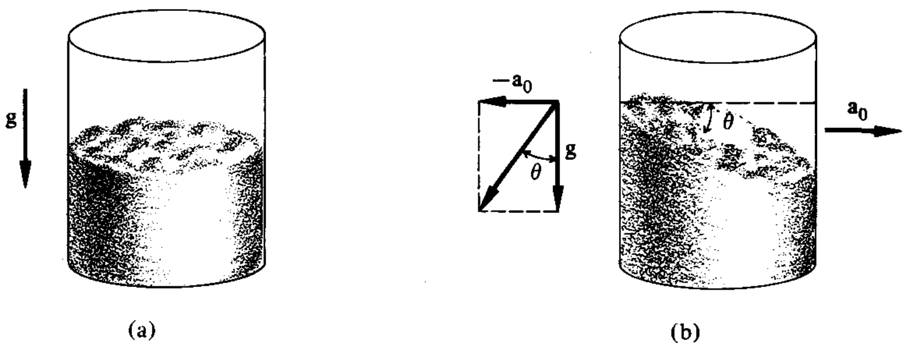


Figure 14-6

Correspondingly, Figure 14-6b shows the situation when the system undergoes an acceleration $\mathbf{a}_0$ directed to the right. Again, the surface of the liquid will be at right angles to the effective force of gravity on each molecule of the liquid. However, now this force is the vector sum of the gravitational force $m\mathbf{g}$ acting on a molecule of mass m and a "fictitious" force of strength ma_0 acting horizontally and to the left in the figure (see (4-26)). In accordance with the vector diagram in the figure, the net force acting on each molecule has the strength $m(g^2 + a_0^2)^{1/2}$ and makes an angle θ with the vertical given by

$$\tan\theta = \frac{a_0}{g}$$

Since the surface of the liquid will be at right angles to the direction of this force, it follows from the geometry in the figure that θ is also the sought-for angle between the surface of the liquid and the horizontal.

14-4 Archimedes' principle

Consider, in Figure 14-7, a rectangularly shaped body of thickness h and of cross-sectional area A submerged below the surface of a liquid of density ρ. If

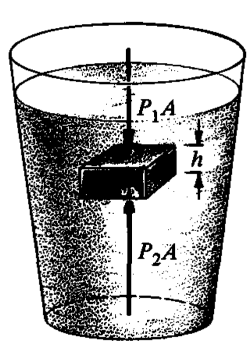


Figure 14-7

P_1 and P_2 are the liquid pressures at the top and bottom of the body, respectively, then the net upward force of the body due to liquid pressure is $(P_2 - P_1)A$. According to (14-5), this pressure difference is ρgh. Hence the net upward force on the body due to the pressure of the liquid is

$$(P_2 - P_1)A = \rho ghA$$

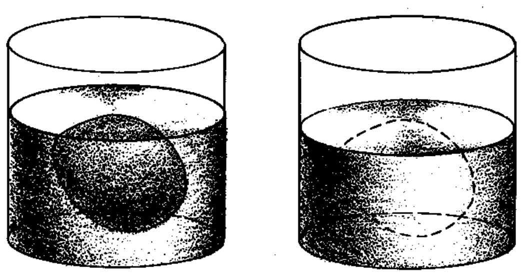
But hA is the volume of the body, and thus the right-hand side of this relation represents the weight of the *liquid displaced by the body*. Hence we see that, at least for the situation shown in the figure, a body submerged in a liquid experiences an upward force equal to the weight of the liquid it displaces.

This principle, which is more generally valid than is implied by this derivation, is known as *Archimedes' principle*. It may be stated formally as follows:

A body wholly or partly submerged in a liquid is buoyed up by a force equal to the weight of the liquid displaced.

Although stated in terms of liquids this principle is equally valid for bodies submerged in gases. Because of the greater density of liquids, however, the effect is much more pronounced for bodies submerged in liquids.

We may justify the principle of Archimedes for arbitrarily shaped bodies in the following way. Consider, in Figure 14-8a, a body submerged below the surface of a liquid. To establish that the buoyant force on it is precisely the same as is the weight of the displaced liquid, consider in Figure 14-8b the same situation as before, but with the body replaced by an equal volume of liquid.



(a) (b)

Figure 14-8

Let us now focus attention on the particular portion of the liquid that has precisely the same shape as does the body and is represented by the dashed surface in the figure. Since this portion of the fluid is in mechanical equilibrium, the net force acting on it due to the pressure of the surrounding fluid must be numerically equal to its own weight. If we now replace this part of the fluid by a body of precisely the same shape, the pressure due to the surrounding liquid molecules will not be altered and thus the body will

experience an upward buoyant force equal to the weight of the liquid it displaces. The validity of Archimedes' principle is thereby established.

A similar argument shows that the principle is also valid for a body only partially submerged in a fluid, such as a piece of wood floating on the surface of a lake, for example.

In order to illustrate the meaning of this principle more fully, consider in Figure 14-9 an object of weight W and volume V suspended from a spring scale above a beaker containing a homogeneous incompressible liquid of density ρ_0. Figure 14-9a shows the situation in which the object is not submerged in the liquid; here the reading on the scale (neglecting the buoyant force of the air) is the total weight of the object W. As shown in the force diagram to the left of the figure, the body is subject to a downward gravitational force W and an upward force T due to the spring, and since it is at rest it follows that $T - W = 0$.

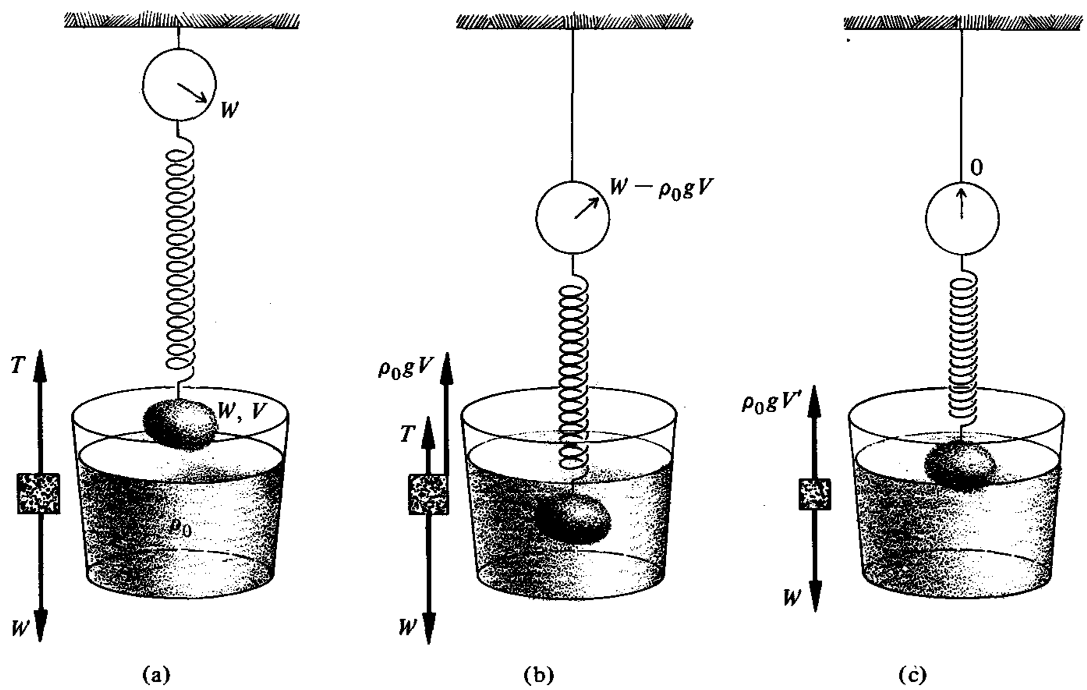


Figure 14-9

Figure 14-9b depicts the case where the weight W of the object exceeds that of an equivalent volume of the liquid, so that $W > \rho_0 g V$. Here the body becomes wholly submerged in the liquid when it is lowered. This time, in addition to an upward force T due to the spring and a downward gravitational force W, there is an upward buoyant force produced by the differential pressure in the liquid. According to the principle of Archimedes, this buoyant force must be equal to the weight $\rho_0 g V$ of the displaced liquid, with ρ_0 the density of the *liquid* and V the volume of the *body*. Applying the conditions for equilibrium, we find that

$$T + \rho_0 g V - W = 0$$

or, equivalently,

$$T = W - \rho_0 g V$$

As shown in the figure, the reading on the spring balance is now $(W - \rho_0 g V)$.

Finally, Figure 14-9c shows the situation in which the weight of the object is less than that of an equivalent volume of the fluid; that is, the case for which $W < \rho_0 g V$. This time, only a volume $V'(< V)$ of the body becomes submerged in the liquid, so the upward buoyant force $\rho_0 g V'$ is precisely equal to the weight of the body. As shown in the figure, the body is here entirely supported by the liquid, and thus it floats. There is no need for the spring to exert a force to support it, and accordingly the spring balance reads zero in this case.

Example 14-4 A meteorological balloon of mass 5 kg has a volume of 10 cubic meters and is filled with 2 kg of helium. Assume that the density of air at sea level is $1.3\ \text{kg/m}^3$ (see Table 13-1).

(a) Calculate the buoyant force on the balloon.
(b) What is the upward acceleration of the balloon?
(c) Discuss qualitatively how high the balloon will rise.

Solution

(a) According to Archimedes' principle, the upward buoyant force on the balloon is equal to the weight W_0 of the air displaced. Using the given data, we have

$$\begin{aligned} W_0 &= \rho_0 g V = (1.3\ \text{kg/m}^3) \times (9.8\ \text{m/s}^2) \times 10\ \text{m}^3 \\ &= 1.3 \times 10^2\ \text{N} \end{aligned}$$

(b) The total weight of the helium and the balloon is

$$(m_1 + m_2)g = (5\ \text{kg} + 2\ \text{kg}) \times 9.8\ \text{m/s}^2 = 69\ \text{N}$$

and it follows, by use of (a), that the net upward force is

$$1.3 \times 10^2\ \text{N} - 69\ \text{N} = 61\ \text{N}$$

Accordingly, the acceleration a_0 of the helium and the balloon is given by the ratio of this force to the mass, 7 kg; that is,

$$a_0 = \frac{61\ \text{N}}{7\ \text{kg}} = 8.7\ \text{m/s}^2$$

(c) Because of the fact that the density of the air decreases with elevation, the weight of the displaced air also decreases gradually as the balloon rises. Eventually, the balloon reaches an elevation where the weight of the displaced air is precisely 69 newtons; here its ascent stops.

Example 14-5 A certain volume V of an iron sphere (mass density $\rho_{\text{Fe}} = 7.8 \times 10^3\ \text{kg/m}^3$) floats in a pool of mercury (mass density, $\rho_{\text{Hg}} = 13.6 \times 10^3\ \text{kg/m}^3$). What fraction of the iron is submerged under the surface of the mercury?

Solution If V' is the volume of the submerged iron sphere, then the upward buoyant force on it, according to Archimedes' principle, is $\rho_{\text{Hg}} g V'$, since V' is the volume of the

mercury displaced. The total weight W of the iron ball is $\rho_{Fe} g V$. Applying the conditions for equilibrium, we have

$$\rho_{Hg} g V' - \rho_{Fe} g V = 0$$

and this leads to the desired volume ratio

$$\frac{V'}{V} = \frac{\rho_{Fe}}{\rho_{Hg}} = \frac{7.8 \times 10^3 \text{ kg/m}^3}{13.6 \times 10^3 \text{ kg/m}^3} = 0.57$$

Example 14-6 A small body of mass density ρ is dropped from a height h into a lake of density ρ_0 ($> \rho$), as in Figure 14-10. Neglecting all frictional and other dissipative effects:

(a) What is the velocity of the body just prior to entering the water?
(b) What is the acceleration of the body while it is under water?
(c) What is the maximum depth x to which the body will sink prior to returning to float on the surface?

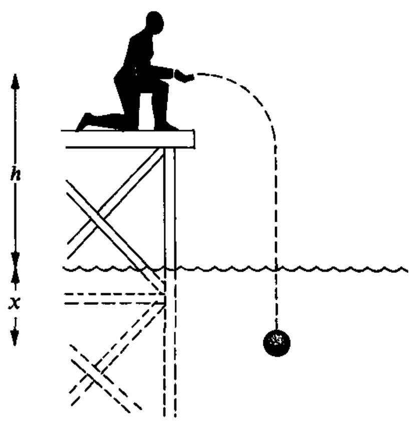


Figure 14-10

Solution

(a) Assuming that the body is dropped from rest, we find by use of (2-18) that its velocity v_0 just prior to entering the water is determined by

$$v_0^2 = 2gh$$

(b) Once the body is submerged under water, it will experience the downward gravitational force $\rho g V$ and an upward buoyant force due to liquid pressure $\rho_0 g V$, where V is the volume of the body. Thus there is a net upward force $gV(\rho_0 - \rho)$ acting on the body. Its mass is ρV so that, in accordance with Newton's second law, the upward acceleration a of the body is

$$a = \frac{gV(\rho_0 - \rho)}{\rho V}$$
$$= g\frac{(\rho_0 - \rho)}{\rho}$$

(c) The maximum depth to which the body sinks may be obtained by setting the final velocity v_f in the relation $v_f^2 = v_0^2 - 2ax$ to zero. Using the value for the initial velocity

v_0 as given in (a), we find that

$$x = h\frac{\rho}{\rho_0 - \rho}$$

Note that the density ρ of the body has been assumed to be less than that of water, ρ_0. If the density of the body is greater than that of the liquid, the net acceleration of the body is directed downward, and for this case it comes to rest at the bottom of the lake.

Note also that conditions here have been assumed to be very much idealized. In fact the body will not descend all the way to the depth x calculated above; some of its initial energy will be dissipated, for example, to produce water waves.

14-5 The measurement of pressure

The purpose of this section is to examine some practical questions associated with the measurement of pressure in both liquids and gases.

Consider first, in Figure 14-11, a U-tube containing a liquid of density ρ. Since both ends of the tube are open and the pressure at the surface of the liquid in each arm has the same value P_0, it follows that the height of the liquid in each arm will be the same. To see this, consider the situation at an imaginary surface A at the bottom of the tube. According to (14-4), the pressure on the surface due to the liquid in the left-hand arm is $(P_0 + \rho g l_0)$. Since this must be balanced by a pressure of precisely the same magnitude due to the liquid in the right-hand arm, we see that the height of the liquid in the right-hand arm must also be l_0. In other words, the level of the liquid in the two arms of the U-tube must be the same.

Making use of this type of an argument let us examine the principle underlying the *open-tube manometer,* which is a device that can be used to measure the pressure of a gas. Consider, in Figure 14-12, a U-tube containing a liquid of density ρ and with its right-hand arm exposed to a known pressure P_0 (such as that of the atmosphere). This time, however, instead of exposing the

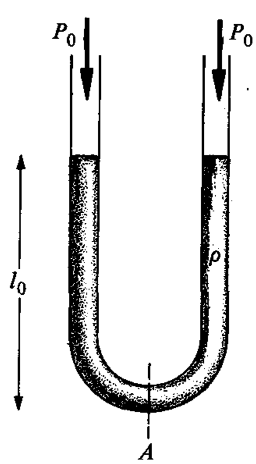


Figure 14-11

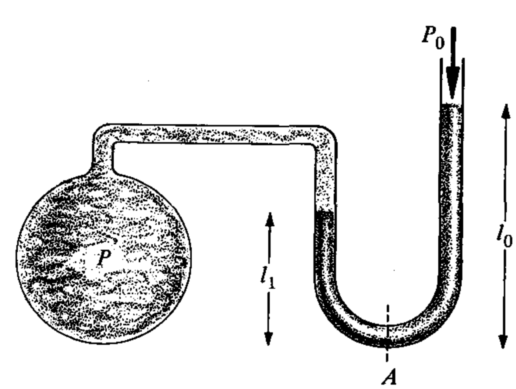


Figure 14-12

left arm of the U-tube to the atmosphere, let us connect it to a container in which is confined the gas whose pressure P we desire to measure. To analyze the system let us proceed, as above, and consider the situation along an imaginary surface A at the bottom of the U-tube. The pressure at A due to the liquid and the gas in the left-hand arm is $(P + \rho g l_1)$, since the pressure at the top of the liquid column is the unknown gas pressure P. In the same way, since the external pressure at the top of the column of liquid in the right-hand arm is P_0, the pressure at A due to the liquid in the right-hand arm is $(P_0 + \rho g l_0)$. Equating these two pressures, we obtain

$$P + \rho g l_1 = P_0 + \rho g l_0$$

or

$$P - P_0 = \rho g (l_0 - l_1) \tag{14-6}$$

Since the quantities ρ, l_0, and l_1 are each easily measured independently, it follows that this device may be used to measure the *gauge pressure* $(P - P_0)$ of the gas. In words, (14-6) says that the gauge pressure of the gas is proportional to the difference in height between the two columns of the liquid in the U-tube. Thus there is a one-to-one correspondence between the difference in height $(l_0 - l_1)$ of the liquid in the U-tube and the gauge pressure $(P - P_0)$ of the gas.

There are a variety of other devices that can be used to measure gas pressure and are based on the same principle as the open-tube manometer. All of these measure the gauge pressure and not the absolute pressure. The actual or absolute pressure of a gas is obtained by adding the gauge pressure to that of the atmosphere.

In the seventeenth century, Evangelista Torricelli (1608–1646) invented a simple means for measuring the (absolute) pressure of the atmosphere. The device he proposed is but a slight variation of the open-tube manometer described above. It is known as a *barometer*, and consists of a long glass tube that is filled with mercury and then inverted into a dish of mercury (see Figure 14-13). As shown, some of the mercury will flow out of the tube and leave behind an evacuated region at the top of this inverted tube. Although some of the mercury atoms will evaporate into the evacuated region above the mercury column, their number is generally not large enough to exert a significant pressure. In the following, we shall assume this region to be a perfect vacuum so that the final result is a column of mercury of a certain height l_0.

Now assuming that this barometer is open to the atmosphere, the pressure on the upper surface of the mercury in the dish is that of the atmosphere P_0. Moreover, since the height of the column of mercury above the surface is l_0, this mercury exerts a pressure $\rho g l_0$ at point A in the figure. Therefore

$$P_0 = \rho g l_0 \tag{14-7}$$

where ρ is the density of liquid mercury, which under the conditions in Table 13-1 has the value 13.6×10^3 kg/m^3, and l_0 is the easily measured height of the mercury in the inverted tube. The atmospheric pressure P_0 may therefore be

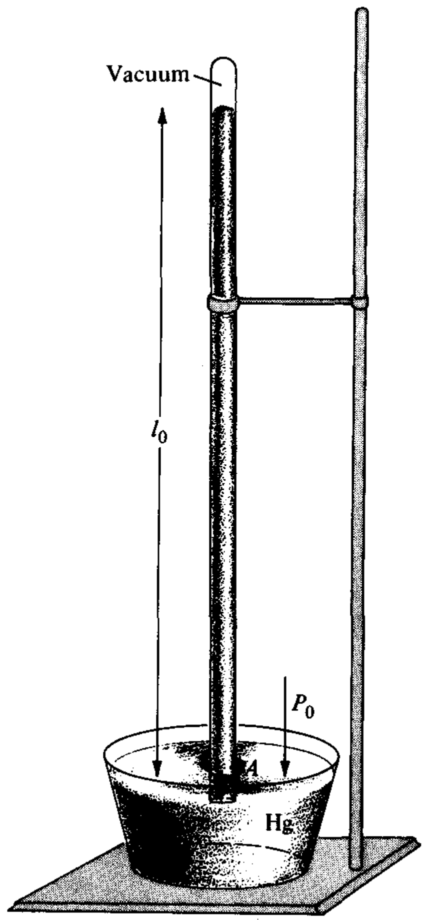


Figure 14-13

ascertained in terms of these two independently measurable parameters ρ and l_0. Experimentally, we find that at sea level the observed value for l_0 is approximately 76 cm.

Because of this fairly simple way of relating atmospheric pressure to the height of a column of mercury, it has been found useful to define a unit of pressure called the *atmosphere.* This unit is defined to be the pressure exerted by a column of mercury of height 76 cm at 0°C at a place on earth where the acceleration of gravity has the value $g = 9.8067\ \mathrm{m/s^2}$. Using the value $\rho = 13.60 \times 10^3\ \mathrm{kg/m^3}$ for the density of mercury at this temperature we find, from (14-7), that

$$\begin{aligned} 1\ \mathrm{atm} &= \rho g l_0 = (13.60 \times 10^3\ \mathrm{kg/m^3})(9.8067\ \mathrm{m/s^2})(0.76\ \mathrm{m}) \\ &= 1.013 \times 10^5\ \mathrm{N/m^2} \end{aligned} \tag{14-8}$$

For most practical purposes it is satisfactory to use the value $1.0 \times 10^5\ \mathrm{N/m^2}$ for this unit of pressure.

On occasion the height of a column of mercury is itself used as a unit of pressure. In this language a pressure of "76 cm" is to be interpreted to mean a pressure of 1 atm $= 1.0 \times 10^5\ \mathrm{N/m^2}$, while a pressure of 38 cm means 0.5 atm.

Example 14-7 What would be the height l_0 of the liquid in the inverted tube in Figure 14-13, if water were used instead of mercury?

Solution Solving (14-7) for l_0 and using the known values $\rho = 10^3\ \text{kg/m}^3$ for water, $g = 9.8\ \text{m/s}^2$, and $P_0 = 1.0 \times 10^5\ \text{N/m}^2$, we obtain

$$l_0 = \frac{1.0 \times 10^5\ \text{N/m}^2}{(10^3\ \text{kg/m}^2)(9.8\ \text{m/s}^2)}$$
$$= 10\ \text{m}$$

This corresponds to a column of water more than 30 ft high!

Example 14-8 Consider, in Figure 14-14, a total length L of a liquid of density ρ in a U-tube of cross-sectional area A. Suppose that the liquid on one side is pushed down a small amount h and then released. Using the ideas of energy conservation, calculate the period of the resulting motion.

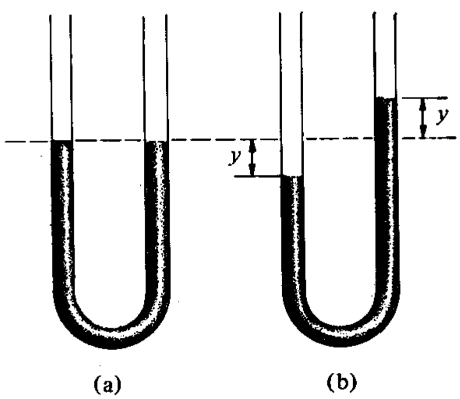


Figure 14-14

Solution Let us define, as in Figure 14-14b, the variable y to represent the displacement of the surface of the liquid from its equilibrium height at any time t. It is apparent that this single variable y uniquely characterizes the motion of the liquid at any time and that at $t = 0$, $y = h$.

The speed of any part of the liquid at any time t is dy/dt. Hence, since the total mass of the liquid is ρAL, the kinetic energy KE of this liquid is given by

$$KE = \frac{1}{2}\rho AL\left(\frac{dy}{dt}\right)^2$$

To calculate the potential energy of the liquid in the gravitational field of the earth, consider in Figure 14-14b the situation in which the liquid is displaced from its equilibrium level by the amount y. This displacement can be thought of as achieved by starting with the equilibrium situation in Figure 14-14a, taking a volume yA of the liquid from one arm of the tube and adding it to the other. In effect, then, a volume of liquid Ay is raised through a vertical height y. The increase of the potential energy of the liquid in this process is

$$V = (\rho Ay)gy = \rho Agy^2$$

Let us now assume that all dissipative effects are negligible, and that thus the total energy of the liquid is conserved. Then the quantity

$$E = KE + V = \frac{1}{2}\rho AL\left(\frac{dy}{dt}\right)^2 + \rho gAy^2$$

where E is the constant total energy of the liquid, is a constant of the motion. Comparing this formula with the corresponding one for a simple harmonic oscillator,

$$E = \frac{1}{2}m\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}kx^2$$

and noting that the period associated with the motion of a harmonic oscillator is $2\pi(m/k)^{1/2}$, it follows that the period of oscillation of the liquid in the tube is

$$2\pi\left(\frac{\rho AL}{2\rho gA}\right)^{1/2} = 2\pi\left(\frac{L}{2g}\right)^{1/2}$$

Thus the period of this motion depends only on the length L of the liquid in the tube and is independent of its density ρ and the cross-sectional area A of the tube.

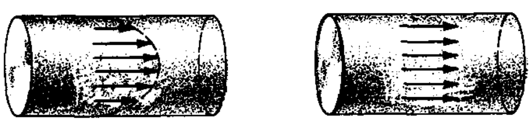
14-6 Fluids in motion

In studying pressure in fluids, we have up to this point been concerned exclusively with fluids at rest. This is the subject matter of *hydrostatics.* For the remainder of this chapter we shall examine briefly the neighboring discipline of *hydrodynamics*; this deals with liquids and gases in motion.

In hydrostatics, a fluid is characterized in terms of the pressure P and the density ρ at each point. In this connection, (14-3) was the fundamental relation between these two physical quantities. By contrast, in order to characterize the state of a fluid in motion, in addition to its pressure and density, it is necessary to specify the velocity $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$ at each point $\mathbf{r}$ of the fluid. This quantity $\mathbf{v}(\mathbf{r}, t)$ is known as the *velocity field* of the fluid and represents the velocity of that element of the fluid which at time t is at the space point $\mathbf{r}$. Thus, the three physical quantities in terms of which we characterize the state of a fluid in motion are its pressure P, its density ρ, and its velocity field $\mathbf{v}$. In general, these quantities vary in time, and their dynamical behavior is governed by a set of equations that are analogous to Newton's laws of motion in particle mechanics. (Technically, they are known as the *Navier–Stokes* equations.) As one might expect, the mathematical problems associated with solving these hydrodynamical equations of motion are very great, and in the following we shall confine ourselves to illustrating some of the simplest, but still interesting, features of fluid motion.

To be specific, let us consider the very simple type of fluid motion that technically is known as *steady, incompressible, nonviscous,* and *irrotational.* By the term *steady,* we mean that the velocity field $\mathbf{v}$ of the fluid is independent of time. Physically speaking, in steady flow each element of the fluid passing

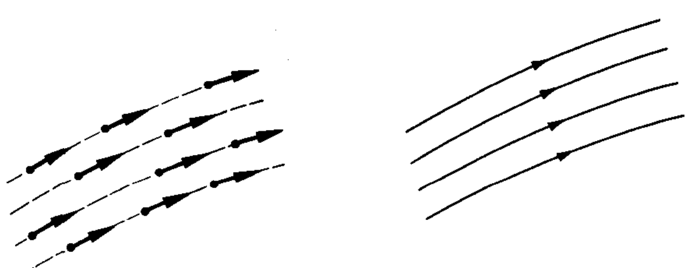
through a given point does so at the same velocity. The term *incompressible*, as previously defined, means that all variations in the fluid density ρ can be neglected. By the term *nonviscous* we mean that there are no frictional or dissipative forces of any type acting on any element of the fluid. By contrast, viscous flow is characterized by the existence of tangential, velocity-dependent forces, which act between layers of a fluid in relative motion. Figure 14-15 illustrates this distinction between viscous and nonviscous flow by reference to fluid flow in a pipe. Figure 14-15a shows the case of viscous flow; here the velocity of the fluid (as represented by the arrows) is a maximum at the center of the pipe and decreases to zero at the edge. Figure 14-15b illustrates the situation for nonviscous flow. Here the velocity of each element of the fluid has the same value across the pipe. Finally, the term *irrotational* means that the velocity field **v** can be related to a potential function in the same way that a conservative force acting on a particle is related to its potential energy (see (8-13)). Irrotational fluid motion is characterized by the absence of eddies or whirlpools. If, for example, a small paddle wheel is placed in a fluid undergoing irrotational motion, it will not rotate as it would if placed in a suitable way in a fluid with whirlpools. Strictly speaking, no fluid motion can ever really be *characterized* as being steady, incompressible, nonviscous, and irrotational. Nevertheless this type of motion represents a useful conceptual idealization and we shall make use of it only as such.



(a) (b)

Figure 14-15

Consider a fluid undergoing this idealized type of motion. As shown in Figure 14-16a, let us draw at each point **r** of the fluid a unit vector along the direction of motion of the fluid at that point. Thus each of the unit vectors in the figure is parallel to the velocity field **v** of the fluid. Imagine drawing a series of continuous, directed, tangent lines, as in Figure 14-16b, so that each of the unit vectors is tangent to one, and only one, of them. These lines are known as



(a) (b)

Figure 14-16

streamlines. Thus, a streamline is a directed line drawn in a moving fluid so that the tangent at each point of the line points along the direction of the fluid motion at that point. By definition, then, an element of the fluid that at some instant is on a given streamline will always travel along that streamline. If we think of the fluid in terms of its constituent particles, the motion of each of these particles is invariably confined to one of these streamlines.

An important property of streamlines is that no two of them can ever cross. For if they did, this would imply more than one value for the fluid velocity at that point and this is simply not possible.

Figure 14-17 shows a small bundle of streamlines, which are perpendicular to and pass through the outer perimeter of a certain area element A in a fluid.

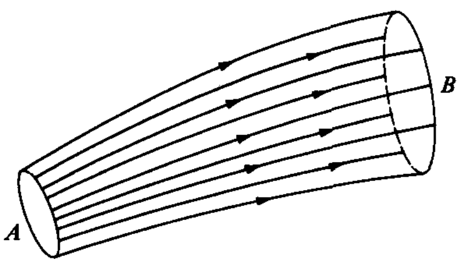


Figure 14-17

We define a *tube of flow*, or a *flow tube*, associated with this area A as the region of space enclosed by these streamlines. Since streamlines can never cross, it follows that no fluid element can ever cross the bounding wall of a flow tube. In other words, all the fluid that enters the flow tube at end A must leave at the opposite end B. In this respect, the steady motion of a fluid through a flow tube is identical to the corresponding flow of a fluid through a pipe whose walls coincide with those of the flow tube. In thinking of a fluid moving along streamlines it is often useful to imagine the flow of a corresponding fluid through a pipe whose walls, at any point, coincide with those of an appropriate flow tube.

14-7 Mass conservation in fluid flow

There are two basic relations that characterize the very simple type of fluid flow with which we are concerned. One of these is known as *Bernoulli's equation* and will be studied in Section 14-8. The second is the law of conservation of mass, which we shall describe now.

Consider, in Figure 14-18, that portion of a flow tube or a pipe between two fixed cross sections of areas A_1 and A_2, respectively. Let us assume that A_1 and A_2 are small enough so that the speed of the fluid is constant across each of them and has the value v_1 across A_1 and v_2 across A_2. During an infinitesimal time interval Δt, the fluid flowing into this region through A_1 has the volume of a cylinder of base A_1 and of height $v_1 \Delta t$. Hence the mass

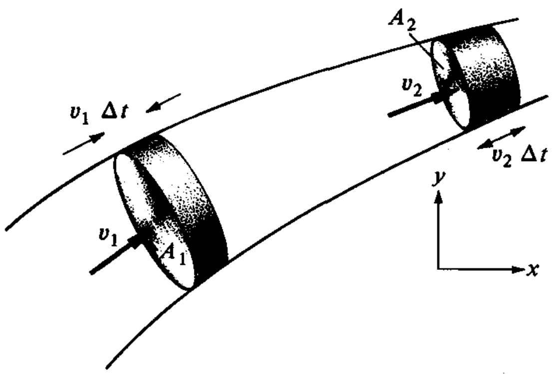


Figure 14-18

flowing into the flow tube at A_1 is $\rho A_1 v_1\ \Delta t$, with ρ the constant fluid density. Correspondingly, the mass flowing out of the flow tube in the same time interval at the other end is $\rho A_2 v_2\ \Delta t$. The net flow of mass Δm into this portion of the flow tube between A_1 and A_2 is thus

$$\Delta m = \rho A_1 v_1 \Delta t - \rho A_2 v_2 \Delta t \tag{14-9}$$

Let us now assume that there are no sources or sinks of fluid in this part of the flow tube. Then, the net mass Δm that flows into this region during any time interval Δt must vanish, and thus, according to (14-9),

$$v_1 A_1 = v_2 A_2 \tag{14-10}$$

Moreover, since the locations of the two faces of the flow tube in Figure 14-18 are arbitrary, we may express (14-10) equivalently as

$$vA = \text{constant} \tag{14-11}$$

In words, this relation says that the product of the velocity of the fluid at any point in a flow tube and the cross-sectional area A at that point is a constant along the entire flow tube.

In physical terms the quantity vA represents the volume of fluid that crosses the area A per unit time. Let us introduce the symbol V_R for this flow rate; that is,

$$V_R = vA \tag{14-12}$$

Then (14-10) and (14-11) tell us that the flow rate V_R is constant along a flow tube. The unit of flow rate is the cubic meter per second (m^3/s). Another unit for V_R is the liter per second, which is defined by

$$1 \text{ liter/s} = 10^{-3}\ \text{m}^3/\text{s}$$

One important consequence of the fact that the flow rate V_R is constant along a flow tube is that the velocity of the fluid is greater the closer together are the streamlines. For as we traverse along a flow tube, as the cross-sectional area A decreases, according to (14-11) the fluid velocity v increases. On the other hand, reference to Figure 14-17 shows that the narrower is the

flow tube, the more tightly together will be the streamlines. To illustrate, consider Figure 14-19, which shows the streamlines associated with the motion of the air about the wing of an airplane. Since the streamlines are closer together above the wing than they are below, it follows that the velocity of the air above the wing is greater than it is below.

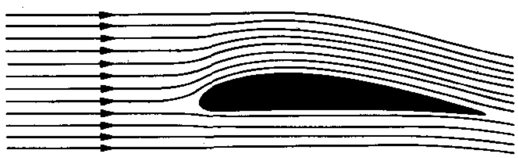


Figure 14-19

A second important consequence of (14-11) is that the pressure of a moving fluid is greater the less its speed. To see this, consider again the situation in Figure 14-18 and assume, as shown, that $A_1 > A_2$. Making use of (14-10), we find that $v_1 < v_2$. In other words, in going from A_1 to A_2 the velocity of the fluid increases. Hence, if there are no external forces such as gravity acting on the fluid, there must be a force due to the fluid pressure that produces this change in velocity. This means that the pressure P_1 at A_1 must be greater than the pressure P_2 at A_2. Hence in the absence of gravity the smaller the speed the greater is the fluid pressure. In Figure 14-19, for example, since the speed above the wing exceeds that below, the fluid pressure is less above the airfoil than it is below. Of course, it is precisely this difference in pressure that produces the lift that airplanes require in order to fly.

Example 14-9 Consider the flow of water at a certain speed v_0 through a cylindrical pipe of radius d. What would be the velocity of this fluid at a point where, because of a constriction in the pipe, the water flow is confined to a cylindrical opening of radius $d/2$?

Solution In terms of (14-10), we are given the data $v_1 = v_0$, $A_1 = \pi d^2$, and $A_2 = \pi(d/2)^2$. Substituting these values, we find that

$$v_2 = 4v_0$$

In other words, since the area of the flow tube, or the pipe in this case, is decreased by a factor of 4, it follows that the fluid velocity must be increased by a factor of 4 in order to be consistent with the conservation of mass.

14-8 Bernoulli's equation

According to (14-4), the pressure at any point in an incompressible fluid at rest depends only on the distance below the surface. For fluids in motion, by contrast, the pressure at any point varies in addition with the speed v of the fluid at that point. The precise form of this variation is given by a relation known as Bernoulli's equation, in honor of its discoverer, Daniel Bernoulli (1700–1782).

Consider again in Figure 14-18 that portion of a fluid confined to a flow tube bounded by the faces A_1 and A_2. Let v_1 and P_1 represent the fluid velocity and the pressure at A_1, and v_2 and P_2 the corresponding quantities at the other end of the flow tube at A_2. To make allowance for the effects of gravity, suppose that the midpoint of A_1 is at a vertical distance y_1 above some horizontal reference line and let y_2 represent the corresponding elevation of the midpoint of A_2. Thus A_2 is at a vertical distance $(y_2 - y_1)$ above A_1.

Let us now calculate the work carried out on this fluid in a certain infinitesimal time interval Δt. The two forces that carry out work on the fluid are (1) the force of gravity and (2) the pressure due to the surrounding liquid. The work carried out by the fluid pressure at A_1 is $P_1 A_1 v_1 \,\Delta t$, since the force $P_1 A_1$ causes a displacement of the fluid of $v_1 \,\Delta t$. Correspondingly, at A_2 the work carried out by the fluid pressure is $-P_2 A_2 v_2 \,\Delta t$, where the minus sign follows since the force here is directed opposite to the motion of the fluid. The work carried out by the force of gravity may be obtained by noting that during this same time interval a mass of fluid $\rho A_1 v_1 \,\Delta t = \rho A_2 v_2 \,\Delta t$ is raised through a vertical distance $(y_2 - y_1)$ corresponding to $-\rho g A_1 v_1 \,\Delta t\,(y_2 - y_1)$ units of work. The total work W carried out on the fluid is the sum of these:

$$\begin{aligned} W &= -\rho g A_1 v_1 \,\Delta t\,(y_2 - y_1) + P_1 A_1 v_1 \,\Delta t - P_2 A_2 v_2 \,\Delta t \\ &= A_1 v_1 \,\Delta t\,[\rho g(y_1 - y_2) + (P_1 - P_2)] \end{aligned} \tag{14-13}$$

where the second equality follows by use of (14-10). In a similar way we may calculate the change in the kinetic energy ΔKE of the liquid by noting that during the time interval Δt a mass $\rho A_2 v_2 \,\Delta t$ of fluid traveling at a speed v_2 leaves that portion of the flow tube of interest while an equal mass $\rho A_1 v_1 \,\Delta t$ traveling at v_1 enters it. Hence the change in the kinetic energy of the fluid is

$$\begin{aligned} \Delta KE &= \frac{1}{2}(\rho A_2 v_2 \,\Delta t) v_2^2 - \frac{1}{2}(\rho A_1 v_1 \,\Delta t) v_1^2 \\ &= \frac{1}{2}\rho A_1 v_1 \,\Delta t\,(v_2^2 - v_1^2) \end{aligned} \tag{14-14}$$

where in the second equality we have made use of (14-10).

Let us now apply the work-energy theorem ($W = \Delta KE$) in (7-22) to the fluid. Substituting the form for the total work W in (14-13) and the change in the kinetic energy in (14-14), we find, on canceling out the common factor $A_1 v_1 \,\Delta t$, that

$$P_1 - P_2 + \rho g(y_1 - y_2) = \frac{1}{2}\rho(v_2^2 - v_1^2)$$

or, equivalently,

$$P_1 + \rho g y_1 + \frac{1}{2}\rho v_1^2 = P_2 + \rho g y_2 + \frac{1}{2}\rho v_2^2 \tag{14-15}$$

Finally, since the subscripts 1 and 2 refer to arbitrary points along the flow

tube, we may state Bernoulli's theorem in the following form:

In steady, incompressible, nonviscous, and irrotational flow the pressure P, the flow of velocity v and the vertical distance y of a fluid element above some horizontal reference line are related to each other along a flow tube by

$$P + \frac{1}{2}\rho v^2 + \rho g y = constant \qquad \textbf{(14-16)}$$

For the special case of a horizontal flow tube, for which the height y is constant, Bernoulli's equation reduces to

$$P + \frac{1}{2}\rho v^2 = \text{constant} \qquad \textbf{(14-17)}$$

This is consistent with our qualitative considerations in the previous section, where we argued that the velocity along a flow tube increases as the pressure decreases, and conversely. A second special case of interest is that for which the fluid undergoes no motion. Here $v = 0$, and Bernoulli's equation becomes

$$P + \rho g y = \text{constant} \qquad \textbf{(14-18)}$$

This is essentially the same as (14-4). If differs only in that the sign before the term $\rho g y$ is positive because of our having selected the positive sense of the y-axis upward.

Example 14-10 Consider, in Figure 14-20, a beaker filled with a fluid of density ρ. Suppose that a very small hole is made in the side of the beaker at a distance h below the surface of the liquid. Assuming that the external pressure is P_0, calculate the velocity v of the fluid as it emerges from the hole.

Solution Let us apply Bernoulli's equation, (14-15), with point 1 at the small opening and point 2 inside the beaker at a distance h below the surface of the liquid. Since, by

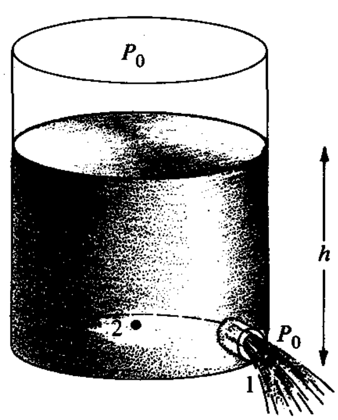


Figure 14-20

hypothesis, the opening at the bottom is very small compared with the cross-sectional area of the beaker, the liquid velocity v_2 at the point 2 is very small and may be neglected. In terms of the parameters in (14-15), we then have the values $P_1 = P_0$, $P_2 = P_0 + \rho gh$, $y_1 = y_2$, $v_2 = 0$, and $v_1 = v$ is the velocity that we wish to calculate. Substituting these values into (14-15), we obtain

$$P_0 + \frac{1}{2}\rho v^2 = P_0 + \rho gh + 0$$

and this simplifies to

$$v = \sqrt{2gh}$$

It is interesting to note that this velocity corresponds precisely to that acquired by a body that starts from rest and falls through a vertical distance h.

Example 14-11 A *Venturi meter* is a device which may be used for measuring the rate of flow of a liquid through a horizontal tube or a pipe. Figure 14-21 shows the essential components of a Venturi meter. A fluid of density ρ_0 travels at a speed v_0—whose value we desire to ascertain—through a pipe of radius R and which has a constriction of radius r. In order to measure the pressure difference between the points 1 and 2 let us connect an open-tube manometer with one end at the constriction and the other at another point of the pipe. If h is the difference in the mercury level in the manometer, calculate the fluid velocity v_0 in the pipe in terms of the remaining parameters.

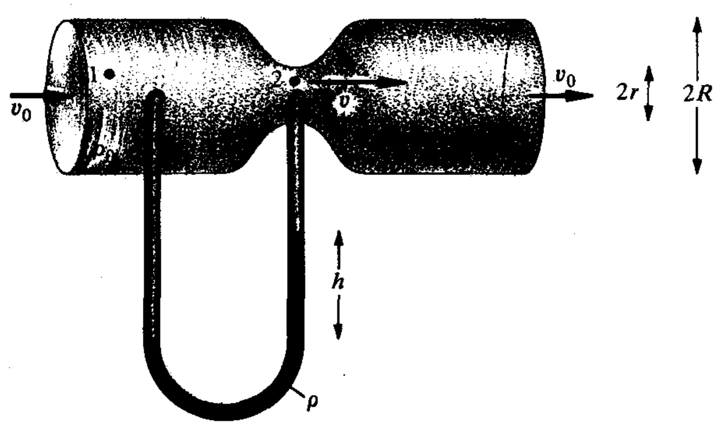


Figure 14-21

Solution If ρ is the density of the mercury, then the measured difference in pressure between the points 1 and 2 is

$$P_1 - P_2 = (\rho - \rho_0)gh$$

where ρ_0 is the density of the fluid in the tube. Let us define v to be the fluid velocity in the constricted part of the tube. Making use of (14-10) and Bernoulli's equation, (14-15), we may then write

$$v\pi r^2 = v_0\pi R^2$$

and

$$P_1 + \frac{1}{2}\rho_0 v_0^2 = P_2 + \frac{1}{2}\rho_0 v^2$$

Finally, eliminating the two quantities $(P_1 - P_2)$ and v and solving the resulting relation for v_0, we obtain

$$v_0 = \left[\frac{2gh(\rho - \rho_0)r^4}{\rho_0(R^4 - r^4)}\right]^{1/2}$$

Since the various quantities ρ, ρ_0, R, and r can be measured independently, it follows that the velocity of flow of the liquid may be ascertained directly in terms of the height differential h.

Example 14-12 Consider, in Figure 14-22, the chamber of a rocket that might be used for propulsion in space. Suppose that the gas in the chamber is maintained at a density ρ_0 and a pressure P_0 and that it escapes into the vacuum through a small opening of area A at one end of the rocket.

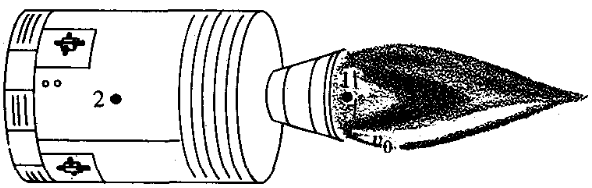


Figure 14-22

(a) Express the exhaust speed v_0 relative to the rocket in terms of remaining parameters.
(b) Calculate the thrust produced on the rocket.

Solution

(a) Assuming that conditions are such that Bernoulli's equation with $g = 0$ applies, we find, by use of (14-5) and the given parameters, that

$$P_0 = \frac{1}{2}\rho_0 v_0^2$$

Here we have used the facts that we are exhausting into a vacuum so that the pressure P_1 outside of the rocket vanishes and that the area A is very small so that the velocity v_2 of the gas inside the chamber may be neglected. Solving for v_0, we find that

$$v_0 = \sqrt{\frac{2P_0}{\rho_0}}$$

(b) To calculate the thrust of the rocket, we note that during a time interval Δt, a mass $\rho_0 A v_0\, \Delta t$ exhausts from the chamber, so that it gains an amount of momentum $(\rho_0 A v_0\, \Delta t)v_0 = \rho_0 A v_0^2\, \Delta t$. Hence the thrust F_{th}, which is the change of momentum per unit time, is

$$F_{th} = \frac{\rho_0 A v_0^2\, \Delta t}{\Delta t} = \rho_0 A v_0^2$$

Making use of the result of (a), we may express this equivalently as

$$F_{th} = 2P_0 A$$

14-9 Summary of important formulas

The pressure $P(y)$ at a distance y below the surface of an incompressible liquid of density ρ is

$$P(y) = \rho g y + P_0 \tag{14-4}$$

where P_0 is the pressure at the surface $y = 0$. Equivalently, we may express this by

$$P_1 - P_2 = \rho g h \tag{14-5}$$

where P_1 and P_2 are the pressures at two points in the fluid separated by a vertical distance h.

For steady, incompressible, nonviscous, and irrotational flow the fact that there are no sources or sinks of matter in a flow tube yields the relation

$$v_1 A_1 = v_2 A_2 \tag{14-10}$$

where v_1 is the fluid velocity at a point in the flow tube with cross-sectional area A_1, and similarly for v_2 and A_2. Bernoulli's equation states that along a flow tube

$$P + \rho g y + \frac{1}{2}\rho v^2 = \text{constant} \tag{14-16}$$

where P is the pressure at a point in the flow tube at a vertical distance y above some reference level and v is the velocity of the fluid at that point.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: (a) incompressible fluid; (b) gauge pressure; (c) absolute pressure; (d) hydrostatics; and (e) barometer.
2. Present an argument to show why the very existence of a liquid implies that the forces between two molecules must have an attractive part. That is, show that if there were no attractive part to the intermolecular force field, then liquids as we know them simply could not exist.
3. In Figure 13-2 we saw that the force between two molecules has in addition to an attractive long-range part, a short-range repulsive part. What properties of liquids that you know of require the existence of such a short-range repulsive force? Could liquids exist if this short-range repulsive part were not present?
4. Describe in both macroscopic and microscopic terms the difference between liquids and gases. Is it possible to state under all circumstances whether a given substance is in a liquid or a gaseous state? Explain.
5. What is the physical mechanism underlying the production of pressure in a liquid? Explain why a molecule of a liquid as it strikes a wall of the confining vessel will, in general, exert a smaller force on it than would a molecule of a gas moving towards the wall at the same incident velocity.
6. Describe in microscopic terms the process of evaporation in a liquid. Why does not all of the liquid in a container exposed to the atmosphere

evaporate instantly? What would happen to a liquid if it were in a vacuum?

7. Consider the slab of area A below the surface of the liquid in Figure 14-2a. Show that it is not possible for the pressure to vary along the horizontal direction by showing that if this were the case, the slab would have to accelerate.
8. Consider again the slab of liquid in Figure 14-2a. Explain in physical terms why the pressure on the upper part of the slab is less than that at the bottom, even though the density of the liquid is the same in both places. In other words, describe in what ways the molecules beneath the slab differ in their activities from those above.
9. Explain in microscopic terms the physical origin for the buoyant force predicted by Archimedes' principle. Consider separately the cases of a dilute gas and of a liquid.
10. Does the buoyant force on an object submerged below the surface of an *incompressible* liquid vary with depth below the surface? What if the liquid were compressible?
11. A body dropped into a liquid starts to sink and comes to rest at a point *above* the bottom of the container. Present an argument to show that this implies that the liquid is not incompressible.
12. A flat block of material lies at the bottom of a tank of water. If there is liquid between the bottom of the block and the tank, the force required to lift it is less than its weight. However, if the bottom of the block is flush against that of the container so that there is no liquid between these two surfaces, then the initial force required to lift the block is very much greater than its weight. Explain.
13. Consider, in Figure 14-23, a long tube AB, which is forced to rotate about a

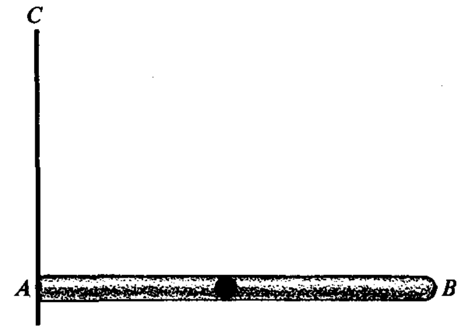


Figure 14-23

vertical axis AC through one end. Suppose the tube contains an incompressible liquid and a small body that is free to move through the liquid. If the body is originally at the center of the tube, which way will it move as the tube rotates? Toward A or B? Consider separately the cases where the mass density of the body is larger and smaller than that of the liquid.
14. What is an open-tube manometer? Explain the purpose for which this device can be used and the principles underlying its operation.
15. Explain what a barometer is and how it can be used to measure absolute pressure.
16. Water is taken out of an open tank by means of a siphon, as shown in Figure 14-24. Explain why the water will flow through the tube in seeming defiance of gravity. For what values of h will the siphon work?

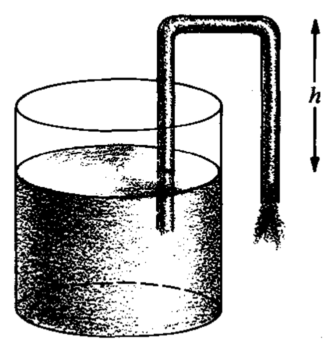


Figure 14-24

17. Define or describe briefly what is meant by the following terms: **(a)** steady flow; **(b)** hydrodynamics; **(c)** irrotational flow; and **(d)** flow tube.
18. Figure 14-25 shows one end of an open-tube manometer exposed to the atmosphere, at pressure P_0, and the other to a wind tunnel along which a gas travels at a certain speed v. Is the pressure in the wind tunnel greater or less than P_0? Can you determine the direction in which the gas moves based only on a knowledge of the difference in height of the mercury in the two arms of the manometer?

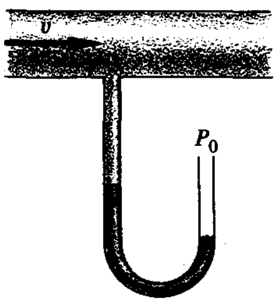


Figure 14-25

19. Using the definition of a streamline, explain why **(a)** streamlines can never cross; **(b)** in regions of higher fluid velocity the streamlines are packed closer together; and **(c)** the pressure is greater the farther apart are the streamlines.
20. Consider the flow of fluid through a constricted portion of a pipe as in Figure 14-21. If P_0 and v_0 are the pressure and velocity of the fluid at point 1, what can you say about the values of the following quantities in the constricted part of the tube at the point 2: **(a)** the pressure P; **(b)** the fluid velocity v; and **(c)** the number of streamlines per unit area going through the pipe at this point.
21. In what way is the motion of a fluid through a flow tube similar to the flow of the fluid through a pipe? Under conditions of steady flow, why must the fluid velocity at the walls of a pipe be directed tangentially to its walls?
22. Wet the end of a piece of paper and let it hang from your lower lip. Explain why the paper will rise as you blow across it.
23. By use of Bernoulli's equation explain in physical terms the mechanism that provides an airplane with lift.

PROBLEMS

1. **(a)** What is the pressure in atmospheres at a point 3 meters below the surface of a lake? What is the gauge pressure there? **(b)** At what depth will the pressure be 2.5 atm?
2. A swimming pool with dimensions of 30 meters × 10 meters is filled to a uniform depth of 2 meters. **(a)** What is the pressure on the floor of the swimming pool? **(b)** What is the total force on the bottom of the swimming pool?
3. A rectangular box of length a and width b is filled to a depth c with a liquid of density ρ. Show that the outward forces due to liquid pressure on two adjacent walls are $\frac{1}{2}\rho gac^2$ and $\frac{1}{2}\rho gbc^2$.
4. Figure 14-26 shows a dam holding back water that is piled up to height h. If the width of dam is a, show that the torque that the water

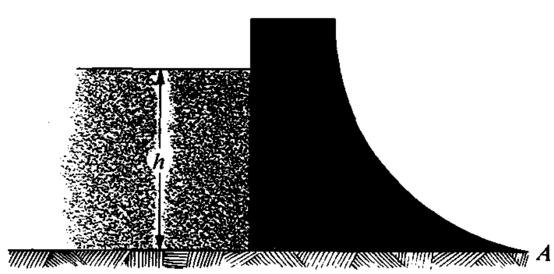


Figure 14-26

pressure produces about the point A in the figure is $\rho gah^3/6$. (*Hint*: Show that the force acting on a horizontal strip of the dam of height dy and at a distance y below the water surface produces about A the torque $\rho agy(h-y)\,dy$.)

5. **(a)** Figure 14-27a shows a cylinder of area A filled to a height h by a fluid of density ρ and kept in place by a force F_1 applied to a piston of area A at the bottom. Calculate the strength of the force F_1 so that the piston remains stationary.

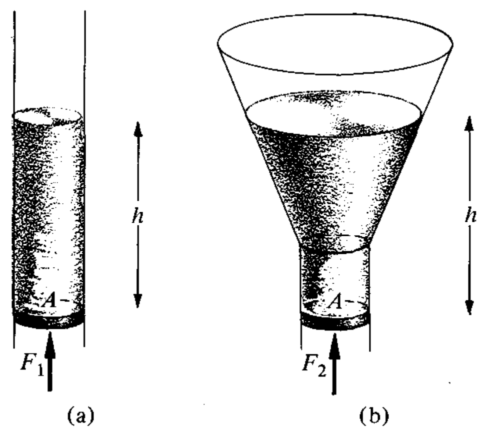


Figure 14-27

(b) Calculate in a similar way the force F_2 that must be applied to the piston in Figure 14-27b to keep the liquid in place.

(c) Compare your answers to (a) and (b) and account for their similarity.

6. A house has a flat horizontal roof 100 square meters in area. What is the total downward force on the roof due to the pressure of the air above it? Explain why, despite the enormity of this force, the roof does not collapse.

7. A liquid of density ρ_0 is poured into one arm of a U-tube that already contains a liquid of density ρ. If the final equilibrium situation is as shown in Figure 14-28, show that the density ρ_0 of the liquid is

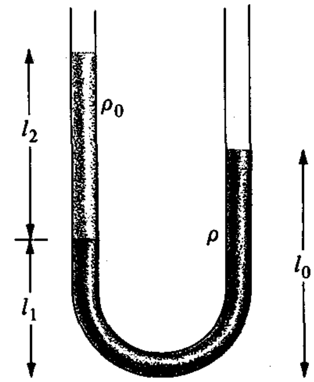


Figure 14-28

$$\rho_0 = \rho\frac{l_0 - l_1}{l_2}$$

8. A man of mass 80 kg is suspended by a rope tied to a horizontal suction cup attached to the ceiling of a room. Assuming that the air trapped between the suction cup and the ceiling has negligible pressure, calculate the minimum area of the cup required to support the man.

9. An evacuated hemisphere of radius a is firmly attached to a vertical wall, as shown in Figure 14-29. If P_0 is the atmospheric pressure, show that the minimum horizontal force F required to pull the hemisphere from the wall is

$$F = \pi a^2 P_0$$

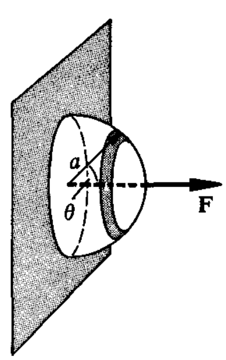


Figure 14-29

(*Hint*: Show that the horizontal force due to the atmosphere on an area element of radius $a \sin\theta$ and thus of area $2\pi a^2 \sin\theta\, d\theta$ of the hemisphere is $2\pi a^2 P_0 \sin\theta \cos\theta\, d\theta$.)

***10.** A solid hemisphere of radius a lies at the bottom of a tank containing water (density ρ) to a height h ($> a$). Assuming that air at atmospheric pressure P_0 is trapped between the flat side of the hemisphere and the bottom of the tank, show that the total initial vertical force required to raise the hemisphere is $\pi a^2 \rho g(h - 2a/3)$. Neglect the weight of the hemisphere.

11. Suppose that the region between two tightly fitting hemispherical shells each of radius 30 cm is evacuated. Making use of your result to Problem 9, calculate the minimum force that must be applied to each hemisphere to pull them apart. When used in this way such hemispheres are known as "Magdeburg hemispheres," in honor of Otto Von Guericke, the Burgermaster of Magdeburg, who in 1654 gave a striking demonstration—involving two teams of horses—of the enormity of this force.

12. If the liquid in the manometer in Figure 14-12 is mercury and the height in the right-hand arm exceeds that in the left by 2 cm, what is the gauge pressure of the gas in the container? What would the pressure in the container have been if the mercury in the left arm were above that in the right arm by 20 cm?

13. If the upper surface of the water in a truck that is traveling along a horizontal road makes an angle of 20° with the horizontal, what is the acceleration of the truck?

14. Suppose that, due to certain changes in the weather, the mercury level in a barometer falls by 0.2 cm. What change in the atmospheric pressure has taken place? What does a rise of 0.2 cm in this level mean?

15. Suppose that the tube of mercury of the barometer in Figure 14-13 has a total length of 90 cm. What is the minimum angle that this tube must make with the vertical so that it is filled entirely with mercury?

16. Suppose that the barometer in Figure 14-13 were placed in an elevator. **(a)** To what height l_0 would the mercury in the tube go when the elevator has an upward acceleration $0.2g$? **(b)** To what height would the mercury in the tube go when the elevator has a downward acceleration of $0.2g$?

17. A rectangular box contains a liquid and accelerates down a smooth inclined plane of angle α, as shown in Figure 14-30.

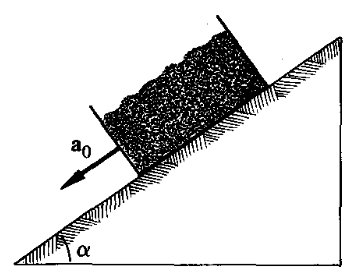


Figure 14-30

(a) Neglecting all effects due to friction, what is its acceleration down the inclined plane?

(b) Show that the upper surface of the liquid in the container will be parallel to that of the inclined plane.

18. A very thin U-tube is filled with water and, as shown in Figure 14-31, has an acceleration $a_0 = 0.5g$, directed as shown. The height of the water in one arm is 10 cm and the distance between the arms is 8 cm.

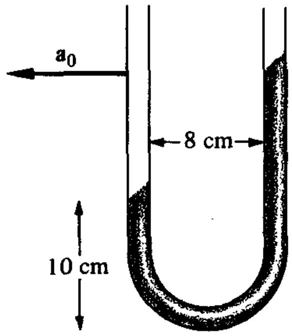


Figure 14-31

(a) Calculate the height of the water in the other arm.
(b) What is the liquid pressure at the lowest point in the tube?

19. Suppose that in Figure 14-23 the tube has a length L, and is filled with an incompressible liquid of density ρ_0 and rotates in a horizontal plane at the angular velocity ω about a vertical axis through one end.
(a) If $P(x)$ is the pressure of the liquid at a point a distance x away from the axis of rotation, show that

$$A_0P(x+dx) - A_0P(x) = (\rho_0 A_0\, dx)\omega^2 x$$

where A_0 is the cross-sectional area of the tube.
(b) If P_0 is the fluid pressure at the origin, $x = 0$, show that the pressure $P(x)$ at the position x in the liquid is

$$P(x) = P_0 + \frac{1}{2}\rho_0\omega^2 x^2$$

*20. Repeat both parts of Problem 19, assuming this time that the tube makes an angle α with the vertical but that the axis of rotation is still vertical and goes through the lower end of the tube.

21. Consider the hydraulic lift illustrated in Figure 14-32, where an incompressible liquid is in a container that is completely enclosed

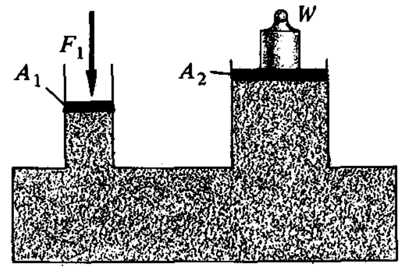


Figure 14-32

except for two movable pistons of respective areas A_1 and A_2.
(a) If a force F_1 is applied to the smaller piston of area A_1, calculate, by use of Pascal's principle, the increase in the liquid pressure everywhere.
(b) Show that the weight W that can be supported on the second piston by this force F_1 is

$$W = \frac{F_1A_2}{A_1}$$

22. Suppose that the hydraulic lift in Figure 14-32, is used to raise the weight W by a distance d_2.
(a) To what distance d_1 must the other piston be depressed?
(b) How much work must be supplied by the agent who exerts the force F_1?
(c) Compare your answer to (b) with the change in potential energy of the weight.

23. Consider again the hydraulic lift in Figure 14-32.
(a) If the area of the smaller piston is 5 cm^2 and that of the larger is 100 cm^2, what force is required at the smaller piston to support a load of 500 kg on the other one?
(b) How far down must the smaller piston be depressed so that the load is raised a distance of 10 cm?

24. A cubical block of material has a volume of 30 cm^3 and floats in water. If two thirds of the volume of the block is below the surface of the

water, calculate the mass of the block and its density.

25. Making use of the fact that the ratio of the densities of seawater to ice is 1.12, show that the fraction of an iceberg that is submerged below the surface of the ocean is 0.89.

26. The radius of a balloon filled with helium gas is 50 meters. Assuming the value 1.3 kg/m^3 for the density of air, calculate the maximum total mass that can be lifted by this balloon.

27. A hollow, spherical ball has a mass of 0.4 kg and is composed of a material having a density of 5.0×10^3 kg/m^3. If a vertical force of half its weight is required to support it when submerged in water, what is the volume of the cavity inside this sphere?

28. What must be the volume of a dry dock that is designed to support ships of mass 10^6 kg? Assume that the weight of the dry dock itself can be neglected.

29. A cubical block of side l has a density ρ and floats in a liquid of density $\rho_0 (> \rho)$.
    - **(a)** What fraction of the volume of the block is below the surface of the liquid?
    - **(b)** What force is required to displace the block downward into the fluid by a small amount h?
    - **(c)** If the block is suddenly released, show that the resultant motion is simple harmonic, with a period of $2\pi(l\rho/g\rho_0)^{1/2}$.

30. A small body of density ρ is at a distance h below the surface of a body of water of density $\rho_0 (> \rho)$. It is suddenly released:
    - **(a)** Show that its upward acceleration while still in the water is $g(\rho_0 - \rho)/\rho$.
    - **(b)** What is its velocity just as it reaches the surface?
    - **(c)** Neglecting all dissipative effects, calculate the height to which the body will rise.

31. A cubical block of side l has a weight W and lies at the bottom of a tank containing a liquid of mass density ρ.
    - **(a)** How much force is required to lift the block if there is initially some liquid between the bottom of the block and the bottom of the tank?
    - **(b)** If the bottom surface of the block is flush against that of the tank so that there is no liquid between these two surfaces, show that initially the force required to lift the block is

    $$W + \rho g l^2(h - l) + P_0 l^2$$

    where P_0 is the atmospheric pressure and h is the depth of the water in the tank.
    - **(c)** Explain in physical terms why the "buoyant" force on the block acts along the direction of gravity in this case, rather than in the opposite direction as usual.

32. A rectangular box has an area A and is filled with water of density ρ_0 to a depth h. If a small body of volume V and density ρ $(< \rho_0)$ is now placed in the water, calculate:
    - **(a)** The rise in the water level.
    - **(b)** The pressure which the water exerts on the lower surface of the container. Give a physical interpretation of this result.

33. What is the fluid velocity v associated with the flow of water that issues from a hose of radius 1.1 cm at the rate of 50 cm^3/s?

34. If the hose in Problem 33 has a constriction back of the opening corresponding to a radius of 1.0 cm, calculate:
    - **(a)** The fluid velocity at the constriction.
    - **(b)** The pressure of the fluid at the

constriction relative to that of the atmosphere (P_0), assuming that the hose is horizontal.

35. Consider, in Figure 14-33, water of density ρ flowing out of the bottom of a funnel, which is in the form of a truncated cone of half-angle α. Assuming that the water comes out at a speed v_0 through an opening of radius r at the bottom, find the fluid velocity $v(y)$ at a height y above the base of the funnel. Calculate the flow rate V_R at the height y, and show explicitly that it is independent of y.

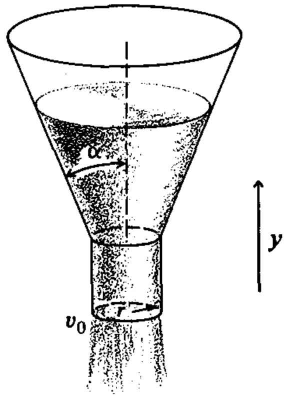


Figure 14-33

36. If the pressure P at some point in a moving fluid becomes very small, then we observe the appearance of bubbles, which are associated with the release of dissolved gases. This phenomenon is known as *cavitation*. Show that if water issues from a horizontal pipe of area A with a velocity v_0, then cavitation will occur if upstream there is a constriction with area a given by

$$a = \frac{A}{[1 + 2P_0/\rho_0 v_0^2]^{1/2}}$$

where ρ_0 is the density of the fluid and P_0 is the pressure of the atmosphere.

37. In Figure 14-20, if the height h of the water in the tank is 1.5 meters and the area A of the small opening at the bottom from which the water escapes has the value $A = 10\text{ cm}^2$ calculate:
(a) The velocity v_0 with which the water flows from the opening.
(b) The flow rate V_R.

*38. Suppose, in Figure 14-20, that the cross-sectional area of the cylinder is A, the area of the small opening at the bottom of the tank is a, and the height of the liquid above the small opening at any time t is $h \equiv h(t)$. Show that the rate dh/dt at which the water level falls is

$$\frac{dh}{dt} = -\sqrt{2gh}\,\frac{a}{A}$$

39. Making use of the result of Problem 38, show that the time τ it takes for the level of the water in the tank to fall from a height h_1 to a height h_2 is

$$\tau = \left(\frac{2}{g}\right)^{1/2} \frac{A}{a}[\sqrt{h_1} - \sqrt{h_2}]$$

40. Consider, in Figure 14-34, a rain barrel standing on a platform of height y_0. If a small hole is punched in the bottom of the barrel, it is found that the resultant stream of water strikes the ground at a distance x_0 from the box.
(a) Show that the velocity v_0 of the

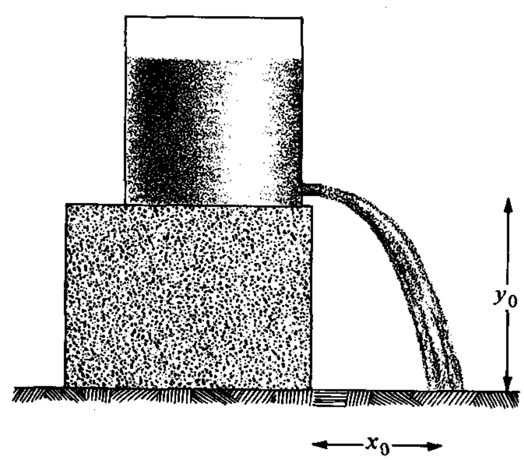


Figure 14-34

water as it comes out of the hole is

$$v_0 = \left[\frac{g{x_0}^2}{2y_0}\right]^{1/2}$$

(b) Calculate the height h of the water level in the barrel above that of the small opening in the bottom.

41. Suppose the liquid flowing through the Venturi meter in Figure 14-21 is water and that the meter itself is characterized by the parameter values $R = 10$ cm and $r = 6$ cm. If the difference in the level of the mercury in the two columns of the manometer is 10 cm, calculate:

(a) The velocity v_0 of the water.

(b) The velocity v of the water in the constricted region.

(c) The flow rate.

42. Suppose that the velocity of the air on the top of an airplane wing is 100 m/s and that on the bottom it is 80 m/s. Assume that the density of the air is 1 kg/m^3.

(a) What is the difference in pressure on the two sides of the wing?

(b) Assuming the plane has a mass of 2×10^3 kg, what must be the minimum area of the wing so that the plane flies?

43. Suppose that as a result of a horizontal stream of air blowing across the top of one of the arms of the U-tube in Figure 14-11, the mercury level in that arm rises by 1 cm. What is the velocity of the airstream? Assume for air the density 1.3 kg/m^3.

15 Temperature and its measurement

For we do not think that we know a thing until we are acquainted with its primary conditions or first principles and have carried our analysis as far as its simplest elements.

ARISTOTLE (***Physics***, Book I)

15-1 Introduction

We continue our study of macroscopic systems in thermal equilibrium by introducing in this chapter a physical quantity that plays a dominant role in the analysis of these systems. This new thermodynamic variable is *temperature*. Qualitatively speaking, temperature is a measure of how hot or how cold a physical system is. Unlike pressure, which was defined only for liquids and gases, the concept of temperature is applicable without exception to *all* thermodynamic systems, including such diverse ones as ferromagnets, metals, stellar interiors, and biological organisms. Because of this all-pervasive utility, the concept of temperature plays an even more important role in the description of systems in thermal equilibrium than does that of pressure.

For reasons of simplicity, in much of the following we shall be concerned with the temperature of a dilute gas. This does not mean that the ideal gas, or any real gas for that matter, is in any way a more important physical system than is any other. But rather, because of its inherent simplicity, the ideal gas

makes it possible not only to define temperature very precisely but also to relate the concept of temperature to the properties of the gas molecules themselves.

15-2 The thermometer and the zeroth law of thermodynamics

A *thermometer* is a device that may be used to define, and thus to measure, the temperature of a physical system. Underlying the validity and the reproducibility of the temperature measurements made by use of a thermometer are certain experimental facts, which are customarily referred to as the *zeroth law of thermodynamics.* This law states:

> *Suppose that two systems, A and B, are in states such that when placed separately into thermal contact with a third system C they are in thermal equilibrium with it. Then A and B will also be in thermal equilibrium with each other if they are placed in thermal contact.*

In other words, any two thermodynamic systems that are in thermal equilibrium with a third system must also be in thermal equilibrium with each other. By the term *thermal equilibrium* we mean a certain quiescent state achieved by a system in which it does not gain or lose energy at the expense of its surroundings. Hence, for example, two systems are in thermal equilibrium with each other if neither gains energy from the other when they are in thermal contact. The term *thermal contact* here means that the systems are physically connected or contiguous in a way so that an exchange of energy between them is possible or at least not prevented.

To illustrate, let system A be a cup of boiling water, system B a cup of ice, and system C the air in an ordinary room. Suppose first that the cup of boiling water is exposed to this relatively cooler air. After enough time has elapsed, the water will cool off and come into thermal equilibrium with the air in the room. At this point it will feel just as cool or warm to the touch as does the air itself. Imagine now a second experiment in which the ice is brought into the same room. Originally, it will feel cold to the touch, but again with the passage of time the ice will melt and the resultant water will warm up and come into thermal equilibrium with the air. At this point it also will feel just as cool or warm to the touch as does the air itself. If now the two cups of water A and B are placed into thermal contact with each other, then, in complete accord with the zeroth law, we find that they are in thermal equilibrium with each other. That is, even though the energies of the water in the two containers may be very much different—for example, because one of them contains much more water than does the other—there is no tendency whatsoever for either one to gain energy at the expense of the other. In other words, the two containers are

in thermal equilibrium with each other, since before being placed in thermal contact they were both in equilibrium with a third system, namely the air in the room.

In addition to feeling different to our sense of touch, there are a variety of other physical attributes of macroscopic systems that are also noticeably affected as they cool off or warm up. For example, if a piece of hot metal is introduced into a cold room, its linear dimensions as well as its electrical resistance decrease measurably as it cools. Similarly, if a balloon filled with hot gas cools off, the pressure of the gas inside the balloon decreases, and as a result the balloon tends to collapse. By making use of any one of these *thermometric properties*, as they are known, we may—subject to certain practical limitations—set up a quantitative measure for temperature by simply correlating the variations in the given property with the degree of hotness or coldness of the system. In this way we may define a temperature scale and its associated thermometer in terms of a fixed thermometric property of any agreed-upon physical system. And once a standard for measuring temperature has been established in this way, the temperature of any other system may be ascertained by allowing the chosen thermometer to come into thermal equilibrium with it. The consistency of such measurements is guaranteed by the zeroth law. For if two distinct thermometers are in thermal equilibrium with a third system, then they also must be in thermal equilibrium with each other and hence consistently give rise to equivalent temperature readings.

To illustrate the principle, let us construct a thermometer by use of the fact that many substances expand when heated. For this purpose, consider a thin piece of wire, say of copper, of fixed cross section. Let l_0 be its length as measured on a given meter stick when, as shown in Figure 15-1a, it is in thermal equilibrium with icewater, and let l_1 represent the corresponding

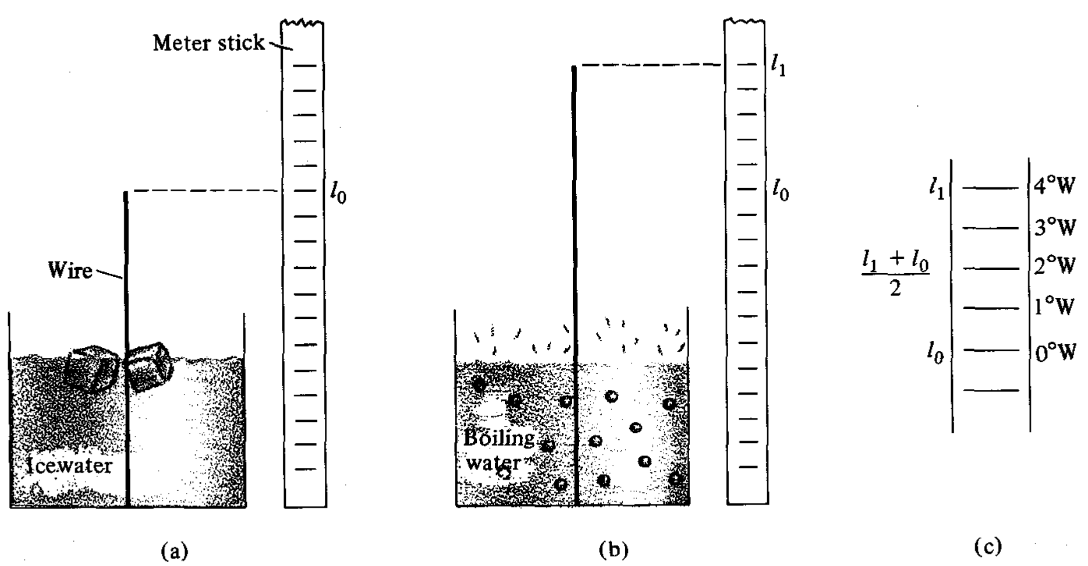


Figure 15-1

length when it is in equilibrium with boiling water at atmospheric pressure. Let us arbitrarily mark off on the meter stick three equally spaced markings between l_0 and l_1. As shown in Figure 15-1c, the meter stick will have markings at the positions l_0, $[l_0+(l_1-l_0)/4]$, $[l_0+(l_1-l_0)/2]$, $[l_0+3(l_1-l_0)/4]$, and l_1.

To define the "W" (for wire) temperature scale let us arbitrarily assign to the lowest marking on the meter stick the value 0°W (read zero degrees W), the next 1°W, and so on, until the mark at l_1 is defined to be 4°W. Let us also extend the scale up and down by making additional marks at intervals of length $(l_1-l_0)/4$ above the mark at l_1 and below that at l_0, respectively. The combination of the given copper wire and the marked meter stick constitutes a unique thermometer. To determine on this scale the temperature of any other system we need only let the wire come into thermal equilibrium with the given system and then measure its length. For example, if when in thermal equilibrium with a certain physical system the wire has a length $[l_0+(l_1-l_0)/2]$, then its temperature is 2°W. Similarly, if the length of the wire were $[l_0+2(l_1-l_0)]$ when in equilibrium with a physical system, then its temperature would be 8°W. For temperatures below that of icewater the length of the wire would be less than l_0. We call such temperatures negative. If, when in equilibrium with a certain system, the wire had the length $[l_0-(l_1-l_0)/4]$, for example, then the temperature of the system would be -1°W. In this way, then, we have available a direct means for correlating temperatures—that is, the degree of hotness or coldness—of any thermodynamic system with the length of the copper wire. Hence it is a thermometer.

15-3 The constant-volume gas thermometer

Among the large number of practical thermometers that have been defined at one time or another there is one that stands out above them all. This is the *constant-volume gas thermometer*. In addition to its practical importance in measuring temperature, the constant-volume gas thermometer has, as we shall see, considerable theoretical importance as well.

To introduce the ideas underlying this thermometer, consider in Figure 15-2 a fixed amount of a dilute gas confined to a certain volume V and at a pressure P as measured by the difference Δl between the mercury levels in the two arms of a U-tube. If ρ_0 is the density of mercury, then the pressure P of the confined gas may be expressed as

$$P = \rho_0 g \, \Delta l \qquad \textbf{(15-1)}$$

so that knowing the values for ρ_0 and g, the pressure P of the gas is directly related to the difference Δl in the mercury levels in the two arms of the tube.

Suppose now that the gas is heated by being brought into thermal contact with a second system. After thermal equilibrium is established, the difference

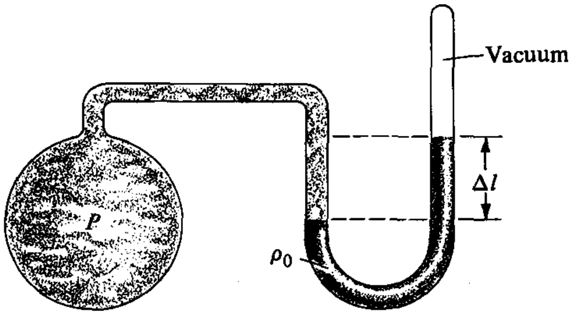


Figure 15-2

in height between the two mercury columns Δl will have increased, thus implying, according to (15-1), that the gas pressure P has also increased. Similarly, if the gas is cooled off, the level difference Δl decreases, since the gas pressure decreases in this cooling process. Hence, just as for the thin-wire thermometer considered in Section 15-2, the correlation between the degree of "hotness" of a physical system and the pressure P of the confined gas can be used to define a temperature scale in terms of this pressure. Note, however, that for the setup in Figure 15-2 as the pressure of the gas changes so does its volume. This implies that any temperature scale defined in terms of this apparatus will depend on the relative linear dimensions of the U-tube and those of the gas container itself. Hence, this particular system is not extensively used in practice.

However, by making a very slight modification of this apparatus we can define a thermometer and an associated temperature scale that are invariant to such geometrical complications. For this purpose it is necessary to arrange things so that as the gas is heated or cooled its volume is kept constant. A simple way for achieving this required constancy of the gas volume is shown schematically in Figure 15-3. This apparatus is a modification of that in Figure 15-2 and involves connecting to the bottom of the U-tube a second flexible tube attached to a mercury reservoir. Since the mercury levels in the two closed arms A and B must always be the same, as the flexible tube is moved up or down, the mercury will flow into or out of the reservoir in order to

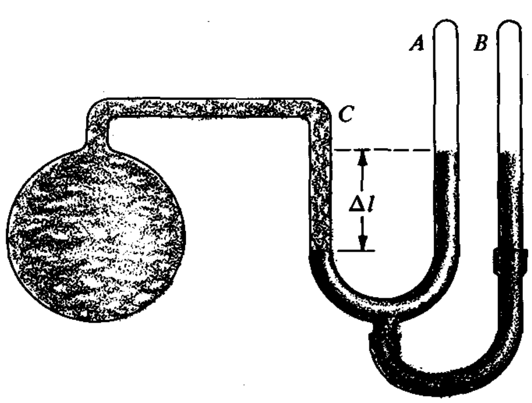


Figure 15-3

maintain this feature. Hence, for any given gas pressure—corresponding to a certain height difference Δl between the mercury levels in arms A and C—we can always move the flexible tube up or down in such a way that the gas occupies any preassigned volume. If heat is added to the gas, for example, then in order to keep the volume fixed, as Δl increases, the reservoir must be moved up by an amount sufficient to keep the mercury level in arm C unchanged. In a similar way, if the confined gas is cooled off, the reservoir must be moved down so that again the gas volume remains fixed even though Δl decreases this time.

15-4 The Kelvin temperature scale

Having available a way for correlating the "degree of hotness" of any thermodynamic system with the pressure of a gas at fixed volume, we shall now define the *absolute* or the *Kelvin* temperature scale.

Let us first define a temperature scale $\bar{T}$ in such a way that the temperature $\bar{T}$ of any thermodynamic system in thermal equilibrium with the gas in Figure 15-3 is proportional to the gas pressure P. That is, we define $\bar{T}$ by

$$\bar{T} = cP \qquad \text{(constant volume)} \tag{15-2}$$

where c is a constant of proportionality and P is the pressure of the gas when it is in equilibrium with the system whose temperature is $\bar{T}$. In operational terms, this means that once the proportionality constant c has been fixed, the temperature $\bar{T}$ of any system is determined by the pressure P of the confined gas in Figure 15-3 when it is in thermal equilibrium with the given system.

To complete the definition of $\bar{T}$ it is necessary to assign a value to the proportionality constant c. Following convention, we shall select this constant by requiring that when a system is in thermal equilibrium with water at its *triple point* (the unique state[1] where the three phases of water, namely ice, steam, and liquid water, are in equilibrium with each other), then its temperature is precisely 273.16. Hence, if P_t is the pressure of the confined *gas* in Figure 15-3 when it is in thermal equilibrium with water at its triple point, then by use of (15-2) we have

$$\frac{P}{P_t} = \frac{\bar{T}}{273.16} \qquad \text{(constant volume)} \tag{15-3}$$

where $\bar{T} \equiv \bar{T}(P)$ is the temperature of the system in thermal equilibrium with the gas when at fixed volume it has a pressure P. For a fixed apparatus of the form in Figure 15-3, P_t will be a unique pressure. Hence, the temperature $\bar{T}$ of *any* physical system may be ascertained by measuring the pressure P of the confined gas when in thermal equilibrium with it and then by making use of (15-3).

[1]See Section 16-9 for a more complete definition.

Unfortunately there is a serious difficulty with this temperature scale $\bar{T}$. For if it is used with different gases or with the same gas but at various densities, slight differences in the corresponding temperatures are found. That is, two different gases that are in thermal equilibrium with each other will not, in general, have *precisely* the same pressure even though they may have the same density. The reason for this is that intermolecular forces, in general, differ for different molecular species, and their effectiveness also varies with gas density. A further complicating feature has to do with the presence of impurities in gases under most circumstances.

It is for these reasons that the *Kelvin temperature scale*, T, is defined by the limit of (15-3) as the particle density n of the gas in the thermometer in Figure 15-3 vanishes. In practice, this means that the gases actually used must be sufficiently dilute that the complications referred to above do not arise. In precise terms, then, the Kelvin temperature scale T is defined in terms of an apparatus of the form in Figure 15-3 by

$$T = [\lim_{n\to 0} \bar{T}] = \left[\lim_{n\to 0} \frac{P}{P_t}\right] 273.16\text{ K} \qquad \text{(constant volume)} \qquad \textbf{(15-4)}$$

where n is the particle density of the gas in the thermometer. The symbol "K" has been introduced for the unit of temperature on this scale.[2] We here anticipate the fact that this unit of temperature of the "kelvin" is identical to that defined on the basis of the absolute or thermodynamic temperature scale, which will be defined in Chapter 17.

It should be noted that the constant-volume gas thermometer as just described cannot, for practical reasons, be used under severe temperature

Table 15-1 Temperatures of various systems and processes

Temperature (K)	*Systems and processes*
0.3	^{3}He boils at low pressure
2.18	^{4}He changes from superfluid to a normal fluid
4.2	^{4}He boils at 1 atm
20	H_2 boils at 1 atm
90	O_2 boils at 1 atm
273.15	Ice melts
273.16	Triple point of water
373.15	Water boils at 1 atm
1234	Ag melts
2485	Ag boils at 1 atm
6000	Surface of the sun
10^5	Surface of O-stars
10^7	Stellar interior
10^8	Thermonuclear reactions occur in plasmas

[2]Prior to 1968 the unit was referred to as the "degree Kelvin," for which "°K" was the symbol.

extremes. A lower limit occurs because of the fact that at sufficiently low temperatures all gases liquefy. In practice this means that its use is restricted to temperatures above a few kelvins. An upper limit occurs naturally enough, since for sufficiently high temperatures the thermometer itself becomes gaseous! The measurement of temperatures that fall beyond these bounds requires the use of other thermometers.

Table 15-1 lists typical temperatures for a number of physical systems. Note the enormous range over which temperatures have a meaning. The fact that all kelvin temperatures are positive follows directly from (15-4). The lowest temperatures attained in the laboratory so far have been obtained by N. Kurti, who has succeeded in achieving temperatures as low as 10^{-6} K.

15-5 Other temperature scales

In applying the laws of thermodynamics to physical systems, it is generally most convenient to express temperatures on the Kelvin scale. In addition, a number of other temperature scales have been defined in the past and some of these are still in current usage. In this section we shall describe three of these scales.

The *absolute Fahrenheit* scale, with the unit of the degree *Rankine* (°R), is defined in the same way as is the Kelvin scale in (15-4), but by assigning the value 491.69 ($= 273.16 \times 9/5$) to the triple point of water instead of the value above of 273.16. If $t_R \equiv t_R(P)$ is the temperature in degrees Rankine when the gas pressure in the constant-volume gas thermometer has the value P, then the analogue of (15-4) in the present case is

$$t_R = \left[\lim_{n \to 0} \frac{P}{P_t}\right] 491.69°\text{R} \qquad \textbf{(15-5)}$$

It is apparent that one degree on the Rankine scale is 9/5 times as large as the degree on the Kelvin scale. Hence, the two scales are related by

$$t_R = 1.8T \qquad \textbf{(15-6)}$$

The Celsius scale[3] is defined by

$$t_C = T - T_0 \qquad \textbf{(15-7)}$$

where $T_0 \equiv 273.15$ and T is the absolute temperature on the Kelvin scale of the thermodynamic system of interest. By definition, the size of the kelvin and the degree Celsius (°C) are identical. On the Celsius scale, for example, the temperature of the triple point of water is 0.01°C, since by definition the absolute temperature T of this system is 273.16 K. A convenient way to remember the Celsius scale is to note that, at a pressure of 1 atm, ice will melt at 0.00°C and water will boil at a temperature of 100.00°C.

[3]Formerly known as the centigrade scale.

Finally we have the *Fahrenheit* scale. If t_C represents the Celsius temperature of any thermodynamic system then the Fahrenheit temperature t_F of that same system is

$$t_F = \frac{9}{5} t_C + 32°\text{F} \tag{15-8}$$

At atmospheric pressure, for example, the freezing and boiling points of water are 32°F and 212°F, respectively. Equivalently, making use of (15-6) and (15-7), we may express t_F in the forms

$$\begin{aligned} t_F &= \frac{9}{5} T - 459.67°\text{F} \\ &= t_R - 459.67°\text{F} \end{aligned} \tag{15-9}$$

The size of the degree on the Fahrenheit scale is precisely the same as that on the Rankine scale, so that a temperature difference of 1°C is the same as that of 1.8°F.

Figure 15-4 shows the relation between these four scales in a graphical way. Consistent with the definition of T in (15-4) and of t_R in (15-5), all temperatures on the Kelvin and the Rankine scales are positive. The lowest point achieved on these two scales has the common value zero and is known as *absolute zero*. The experimental fact that no system can ever attain a temperature of absolute zero, to say nothing of negative Kelvin or Rankine temperatures, is known as the *Nernst heat theorem* or the *third law of thermodynamics*. This law states:

It is not possible by any procedure, no matter how idealized, to cool a system to absolute zero in a finite number of operations.

Note that there is no restriction on how close to absolute zero a system may be taken. We shall establish below that at absolute zero the thermal velocity of a dilute gas vanishes, and this will provide us with an alternate way of understanding why negative Kelvin temperatures cannot be attained.

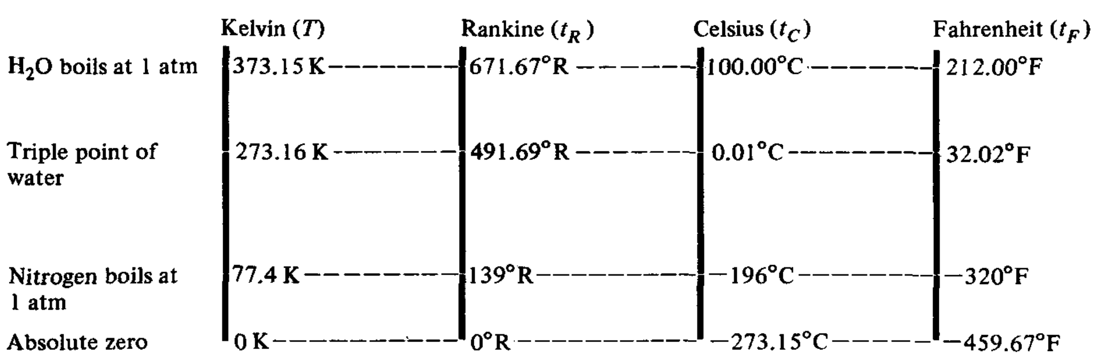


Figure 15-4

Example 15-1 Solid oxygen melts at a temperature of $-218.4°C$. What is this temperature on the other three scales?

Solution Solving (15-7) for the absolute temperature T,

$$T = t_C + 273.15 \text{ K}$$

and substituting the given values $t_C = -218.4°C$, we obtain

$$T = 54.8 \text{ K}$$

Substituting this value for T into (15-6) we find similarly that

$$t_R = 1.8 \times 54.8 \text{ K} = 98.6°R$$

and, finally by use of (15-9), there results

$$t_F = t_R - 459.67°F = 98.6°F - 459.67°F = -361.1°F$$

15-6 The equation of state of the ideal gas

Consider N molecules of a monatomic gas confined to a volume V. Let us suppose that the particle density $n = N/V$ of the gas is sufficiently small so that the inequality, (13-5), is satisfied, and thus that it is an ideal gas. Imagine now carrying out a sequence of measurements of the pressure P of this gas as its temperature T and its density n—or, equivalently, its volume V—are independently varied. Provided that we confine ourselves to low densities and to a range of temperatures for which the gas does not liquefy, the results of all such experiments may be summarized by

$$P = \frac{NkT}{V} = nkT \tag{15-10}$$

In this formula, P represents the pressure of the gas when it occupies a volume V—or, equivalently when it has the density $n = N/V$—and T is its temperature as measured on the Kelvin scale. The symbol k represents a certain empirical constant, known as *Boltzmann's constant*, which has the value

$$k = 1.38 \times 10^{-23} \text{ J/K}$$

Any expression of the type in (15-10) that relates the pressure, volume, and temperature of a gas in thermal equilibrium is known as the *equation of state* for that gas. Hence (15-10) is the equation of state for the ideal gas.

An equivalent, and sometimes more convenient, way of writing (15-10) is

$$PV = \mu RT \tag{15-11}$$

where μ represents the number of moles of the gas

$$\mu = \frac{N}{N_0}$$

and where, in turn, N_0 is Avogadro's number. On comparing (15-10) and (15-11) we see that the parameter R, which is known as the *gas constant*, is the product of Boltzmann's constant k and Avogadro's number N_0. Its numerical value is

$$R = kN_0 = (1.38 \times 10^{-23}\ \text{J/K})(6.02 \times 10^{23}/\text{mole})$$
$$= 8.31\ \text{J/mole-K}$$

For any gas in thermal equilibrium, a plot of pressure as a function of volume, at fixed temperature, is known as an *isotherm* for that gas. It follows from (15-11) that the isotherms for the ideal gas are the hyperbolas PV = constant. Figure 15-5 shows several isotherms. Note that for any two isotherms the one corresponding to the higher temperature lies above the one at the lower temperature. The fact that the isotherms are monotonically decreasing functions is fully in accord with the physical fact that as we increase the pressure on a gas—say by squeezing it—its volume decreases.

The fact that at fixed temperature the product PV of a gas is constant is known as *Boyle's law*. The corresponding constancy of the ratio V/T for a gas whose pressure is held fixed also follows from (15-10), and is known as *Charles' law*.

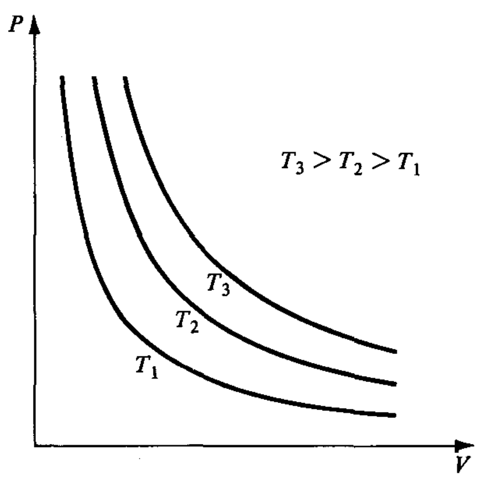


Figure 15-5

Example 15-2 Calculate the volume of 1 mole of an ideal gas at a temperature of 0°C and at a pressure of 1 atm.

Solution According to (15-7), a temperature of 0°C corresponds to the value $T = 273.15$ K. Substituting this value for T and the value $P = 1\ \text{atm} = 1.01 \times 10^5\ \text{N/m}^2$ into (15-11) we find, since $\mu = 1$, that

$$V = \frac{\mu RT}{P} = \frac{1\ \text{mole} \times (8.31\ \text{J/mole-K}) \times (273\ \text{K})}{1.01 \times 10^5\ \text{N/m}^2}$$
$$= 22.4\ \text{liters}$$

Example 15-3 Suppose that at 0°C a certain basketball contains air at a pressure of 1.5 atm. Calculate the pressure of the air in the ball if its temperature is raised to 30°C and its volume increases by 2 percent.

Solution Since the number of moles of gas in the ball is fixed, it follows that μR is constant and we may write (15-11) in the form

$$\mu R = \frac{P_i V_i}{T_i} = \frac{P_f V_f}{T_f}$$

where the subscripts i and f refer to the initial and to the final situations, respectively. Hence

$$P_f = P_i\left(\frac{V_i}{V_f}\right)\left(\frac{T_f}{T_i}\right) = 1.5 \text{ atm} \times 0.98 \times 1.1$$
$$= 1.6 \text{ atm}$$

since, according to the given data, $(V_i/V_f) = 1/1.02 = 0.98$ and $(T_f/T_i) = (273+30)/273 = 1.1$.

Example 15-4 One mole of an ideal gas is originally at a temperature T and at a volume V_0. Suppose it is compressed slowly by an external agent to a final volume $V_0/2$. Calculate the work carried out on the gas and its final temperature if:
(a) The compression is isothermal.
(b) The compression is isobaric.

Solution In general, the work W carried out by an agent during a compression or an expansion is

$$W = -\int P\,dV$$

(a) For an isothermal process the temperature T of the gas is constant. Hence, according to (15-11),

$$W = -\int_{V_0}^{V_0/2} P\,dV = -\int_{V_0}^{V_0/2} RT\frac{dV}{V}$$
$$= -RT \ln V\Big|_{V_0}^{V_0/2} = RT \ln 2$$

and this is represented by the area $ABECDA$ under the T-isotherm in Figure 15-6.

(b) For an isobaric process the pressure P of the gas is kept constant throughout. Hence

$$W = -\int_{V_0}^{V_0/2} P\,dV = -P\int_{V_0}^{V_0/2} dV = -P\left[\frac{V_0}{2} - V_0\right]$$
$$= -\frac{RT}{V_0}\left(-\frac{V_0}{2}\right) = \frac{RT}{2}$$

where the fourth equality follows by the use of (15-11) and the fact that only 1 mole of gas is present. This work is represented by the rectangular area $ADCEA$ in Figure

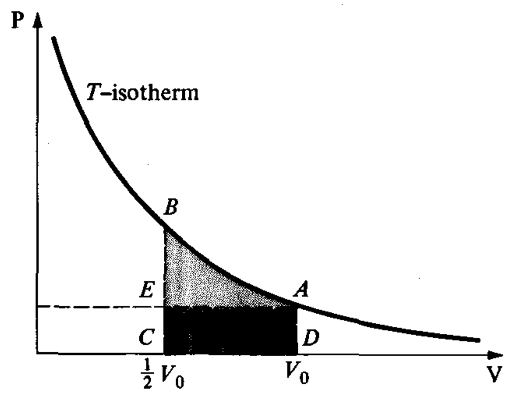


Figure 15-6

15-6. Since P is constant, if follows from (15-10) that T and V are directly proportional to each other. Hence, since the final volume is $V_0/2$, it follows that the final temperature is $T/2$.

15-7 Physical interpretation of temperature for a dilute gas

If two thermodynamic systems at different temperatures are placed in thermal contact, then the one at the higher temperature invariably cools off while the one at the lower temperature warms up. This process will continue until finally they come into thermal equilibrium, at which point they have a common temperature at some intermediate value. Intuitively we expect that in this process of *thermalization*, as it is known, something passes from one of these systems to the other and the question arises as to the nature of this entity that is transmitted. Let us for the moment assign the word *heat* to it. Note that as introduced here the word "heat" is simply a way of referring qualitatively to that something which passes between systems at different temperatures while they are in the process of thermalization. A more precise definition must await our discussion of the first law of thermodynamics in Chapter 16.

According to (13-18), the pressure P of an ideal gas of particle density n may be expressed in the form

$$P = \frac{1}{3} nmv_{\text{th}}^2 \tag{15-12}$$

where m is the mass of a gas molecule and where the thermal velocity v_{th}, as defined in (13-8), is a measure of the average velocity of the molecules in the gas. Substituting this formula for P into the equation of state for the ideal gas in (15-10), we find, on multiplying both sides by $3/2n$, that

$$\frac{1}{2} mv_{\text{th}}^2 = \frac{3}{2} kT \tag{15-13}$$

In words, this simple but very profound result states:

The absolute temperature T of an ideal gas is directly proportional to the square of the thermal velocity; or, equivalently, the absolute temperature of an ideal gas is directly proportional to the average kinetic energy per molecule.

Moreover, since the total energy E of a monatomic[4] ideal gas is the product of the average kinetic energy per molecule and the number N of molecules, (15-13) may be expressed in the equivalent form

$$E = \frac{3}{2} NkT \tag{15-14}$$

Let us now reexamine the question of what we mean by "heat" in light of this very simple but extraordinary profound result. To this end, consider two dilute gases which are at different temperatures and are placed in thermal contact with each other. As we know, the temperature of the cooler one increases, whereas that of the hotter one decreases until eventually they achieve a state of thermal equilibrium at a common intermediate temperature. According to (15-13), a rise in temperature corresponds to an increase in the thermal velocity of a gas and, correspondingly, a decrease of temperature means a decrease in thermal speed. Hence we conclude that what passes between the two gases is nothing material or tangible but rather what is transmitted is a *dynamical characteristic* of the two gases; namely, a *part of the kinetic energy of the hotter gas.*

15-8 Two dilute gases

Consider, in Figure 15-7, two dilute gases A and B, characterized initially by the parameter values m_A, m_B, N_A, N_B, T_A, T_B, v_A°, and v_B° for their respective molecular masses, particle numbers, temperatures, and thermal velocities, and confined to the two compartments of a container. Assume that the external walls of the container are insulated but that the partition between them is *diathermal.* This means that the gases are isolated but can exchange energy freely with each other through the diathermal wall. If initially $T_B > T_A$, then heat flows from B to A until finally thermal equilibrium is established at a certain temperature T lying somewhere between T_A and T_B. In addition to this change in temperature, and the associated change in the thermal velocities of the gases, none of the other gas parameters is altered.

[4]For a gas of diatomic molecules such as oxygen, relations such as that in (15-13) are still valid, provided that we replace the factor "3/2" by 5/2. For simplicity, we omit this complicating feature for the moment and thus our results will be applicable only to monatomic gases.

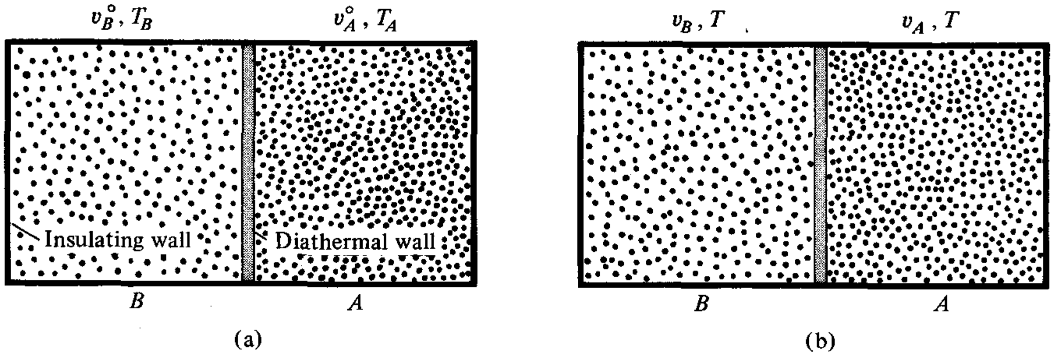


Figure 15-7

Let us now examine what happens to the gas in light of (15-13). According to this formula, the initial parameters v_A°, T_A, v_B°, and T_B are related as follows:

$$\frac{1}{2} m_A v_A^{\circ 2} = \frac{3}{2} kT_A$$
$$\frac{1}{2} m_B v_B^{\circ 2} = \frac{3}{2} kT_B \tag{15-15}$$

Similarly, the common final temperature T and the corresponding final thermal velocities v_A and v_B are related by

$$\frac{1}{2} m_A v_A^2 = \frac{1}{2} m_B v_B^2 = \frac{3}{2} kT \tag{15-16}$$

Now since by hypothesis $T_B > T_A$, it follows from (15-15), assuming that $m_B \cong m_A$, that the initial thermal velocity v_B° is greater than the corresponding thermal velocity v_A°. Hence, when a B molecule strikes an A molecule (via the diathermal wall acting as an intermediary), there is on the average a net transfer of energy from the former to the latter. This implies that the thermal velocity of A tends to increase on the average while that of B tends to decrease. From a microscopic point of view then, the statement that the temperature of A rises while that of B falls means simply that the average kinetic energy of the molecules of A is increasing at the expense of that of B.

A number of important quantitative conclusions also follow from (15-15) and (15-16). First, solving for the ratio v_A and v_B in (15-16) we find that

$$\frac{v_A}{v_B} = \left[\frac{m_B}{m_A}\right]^{1/2} \tag{15-17}$$

or in other words the ratio of the thermal velocities of two gases in thermal equilibrium varies inversely as the square root of their masses. For example, if a dilute gas of H_2 molecules is in thermal equilibrium with a dilute gas of O_2, then since $m(H_2)/m(O_2) = 1/16$, on the average the hydrogen molecules travel at four times the speed of the oxygen molecules. Similarly, since the ratio of the masses of an O_2 molecule to an N_2 molecule is $16/14 = 1.14$, it follows that the ratio of the thermal speeds of N_2 to O_2 in the atmosphere is $\sqrt{1.14} = 1.07$.

Second, making use of (15-14) and the fact that even though the individual systems exchange energy with each other the total energy is conserved, we have

$$\frac{3}{2}kT_AN_A + \frac{3}{2}kT_BN_B = \frac{3}{2}kT(N_A + N_B)$$

Solving this relation for the final temperature T, we obtain

$$T = \frac{N_AT_A + N_BT_B}{N_A + N_B} \tag{15-18}$$

so, as expected, T assumes a value intermediate between T_A and T_B. For example, if the number of A and B molecules is the same, then the final temperature T is the arithmetic average of the initial temperatures $\frac{1}{2}(T_A + T_B)$.

Example 15-5 Calculate the thermal speed of N_2 molecules and of ^4He atoms in air at 0°C.

Solution Solving (15-13) for v_{th} and substituting the value $T = 273$ K and the mass $m = 6.67 \times 10^{-27}$ kg for ^{4}He, we find that

$$\begin{aligned} v_{th}(^4\text{He}) &= \left[\frac{3\,kT}{m}\right]^{1/2} = \left[\frac{3 \times (1.38 \times 10^{-23}\ \text{J/K}) \times 273\ \text{K}}{6.67 \times 10^{-27}\ \text{kg}}\right]^{1/2} \\ &= 1.3 \times 10^3\ \text{m/s} \end{aligned}$$

Similarly, using the value 4.67×10^{-26} kg for the mass of a nitrogen molecule we may calculate its thermal speed. Equivalently, since the ^{4}He atoms and the N_2 molecules are in thermal equilibrium with each other and since their mass ratio is 4/28 it follows from (15-17) that

$$\begin{aligned} v_{th}(\text{N}_2) &= v_{th}(^4\text{He})[m(^4\text{He})/m(\text{N}_2)]^{1/2} = (1.3 \times 10^3\ \text{m/s}) \times [4/28]^{1/2} \\ &= 4.9 \times 10^2\ \text{m/s} \end{aligned}$$

Example 15-6 Half a mole of a monatomic gas is confined to a 10-liter, insulated container and has a temperature of 50°C. If 0.25 mole of the same gas, but at a temperature of 0°C, is added to the same container, what is the final temperature?

Solution Whenever two dilute gases come into thermal equilibrium without loss of energy, (15-18) is valid. Hence, using the value 6.0×10^{23} for Avogadro's number, we have the parameter values $N_A = 3.0 \times 10^{23}$, $N_B = 1.5 \times 10^{23}$, $T_A = 50\ \text{K} + 273\ \text{K} = 323$ K, and $T_B = 273$ K. Substituting into (15-18), we obtain

$$\begin{aligned} T &= \frac{3.0 \times 10^{23} \times 323\ \text{K} + 1.5 \times 10^{23} \times 273\ \text{K}}{3 \times 10^{23} + 1.5 \times 10^{23}} \\ &= 306\ \text{K} \end{aligned}$$

This corresponds to a temperature of 33°C.

15-9 Adiabatic compression of an ideal gas

By an *adiabatic compression* of a gas we mean one in which no energy leaks into or out of the gas while it is in the process of being compressed. A way for achieving an adiabatic compression is shown in Figure 15-8, where a dilute gas is confined to a container with a movable piston. In order that the compression be adiabatic, it is necessary that the piston and the walls of the vessel be of a nature such that on the average all collisions of the molecules with these bounding surfaces are elastic. Hence the only way that the energy of the gas can change is by the collision of molecules with the piston as it moves into the gas. The physical mechanism underlying such a gain or loss of energy due to a moving piston has been described in Example 13-11.

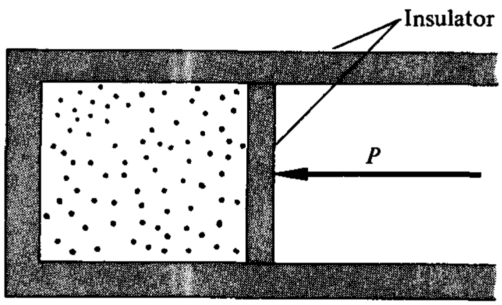


Figure 15-8

Suppose the gas is compressed very slowly, and thus remains in thermal equilibrium throughout. The work dW carried out by the external agent against the pressure P of the gas in changing its volume by an amount dV is $-P\,dV$, according to (13-26). On the other hand, since the process is adiabatic it follows that this work dW must show up as a change of the energy dE of the gas. According to (15-14), if the energy E of a monatomic dilute gas of N molecules changes by an amount dE its temperature T changes by the amount dT given by

$$dE = \frac{3}{2} Nk\,dT \tag{15-19}$$

Hence, equating this change in energy dE to the work dW, we find that for an adiabatic compression

$$\frac{3}{2} Nk\,dT = -P\,dV \tag{15-20}$$

To explore the implications of (15-20), let us substitute, by use of (15-10), for the pressure P on the right-hand side. Dividing the result by NkT, we find that

$$\frac{3}{2}\frac{dT}{T} = -\frac{dV}{V}$$

and this integrates to

$$\ln T^{3/2} = -\ln V + C_1$$

with C_1 a constant of integration. Equivalently, using the property of the logarithm $\ln(xy) = \ln x + \ln y$, we may express this as

$$VT^{3/2} = C \tag{15-21}$$

with C another integration constant. An alternate, and also useful, way for expressing this relation may be obtained by solving (15-10) for T and substituting into (15-21). A brief calculation leads to

$$PV^{5/3} = C' \tag{15-22}$$

where, for a fixed amount of gas, C' is yet another constant.

Suppose, for example, that the pressure, volume, and temperature of an ideal gas are initially P_0, V_0, and T_0, respectively. If the gas undergoes an adiabatic compression, then the corresponding values of these three quantities, P, V, T, are related to the initial values by

$$\begin{aligned} VT^{3/2} &= V_0 T_0^{3/2} \\ PV^{5/3} &= P_0 V_0^{5/3} \end{aligned} \tag{15-23}$$

Figure 15-9 shows a plot of (15-22) for three values of the constant C'. These are known as the *adiabats* for the ideal gas. Also shown on the same P-V diagram are several isotherms (the dashed lines).

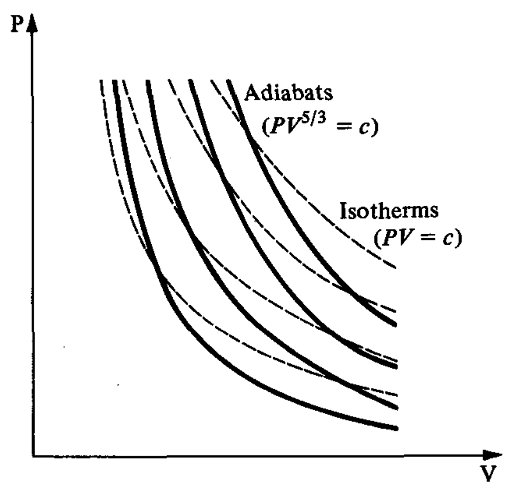


Figure 15-9

Example 15-7 One mole of an ideal gas is subjected to an adiabatic compression as a result of which its volume is halved. If the initial temperature T_0 of the gas is 0°C and its initial pressure is 1 atm, calculate its final temperature T and pressure P.

Solution In terms of the quantities in (15-23), the given parameter values are $V = V_0/2$, $T_0 = 273$ K, and $P_0 = 1$ atm. Solving the first equation of (15-23) for T, we find that

$$T = T_0 \left[\frac{V_0}{V}\right]^{2/3} = 273\ \text{K} \times (2)^{2/3} = 433\ \text{K}$$

Correspondingly, the final pressure P may be obtained by use of the second equation (15-23)

$$P = P_0\left[\frac{V_0}{V}\right]^{5/3} = 1\ \text{atm} \times (2)^{5/3}$$
$$= 3.2\ \text{atm}$$

Note that as the gas is compressed adiabatically, its pressure and temperature both increase, and by a considerable amount in this case.

Example 15-8 Suppose that 1 mole of an ideal gas is subjected to a compression from an initial volume V_0 to a final volume V_f. If P_0 is the initial pressure of the gas, calculate the work W required if:

(a) The compression is isothermal.

(b) The compression is adiabatic.

Solution In general, the work W carried out on a gas in compressing it slowly is

$$W = -\int_{V_0}^{V_f} P\,dV$$

where the choice made for the pressure P as a function of the volume V depends on the particular process involved.

(a) For the case of an isothermal process, according to (15-11) the product PV remains the same throughout the compression. Hence $P = P_0V_0/V$ and, substituting into the formula for W, we find that

$$W = -P_0V_0\int_{V_0}^{V_f}\frac{dV}{V} = -P_0V_0 \ln V\Big|_{V_0}^{V_f}$$
$$= P_0V_0 \ln\frac{V_0}{V_f} \qquad \text{(15-24)}$$

since $\ln 1/x = -\ln x$. For a compression, $V_0 > V_f$ and the work is positive. This is consistent with the fact that in a compression the external agent must carry out positive work on the gas.

(b) For an adiabatic process the pressure P during the compression may be expressed, according to (15-23), by

$$P = P_0\left[\frac{V_0}{V}\right]^{5/3}$$

Hence

$$W = -P_0V_0^{5/3}\int_{V_0}^{V_f}\frac{dV}{V^{5/3}} = -P_0V_0^{5/3}\left[-\frac{3}{2}V^{-2/3}\right]\Big|_{V_0}^{V_f}$$
$$= \frac{3}{2}P_0V_0\left[\left(\frac{V_0}{V_f}\right)^{2/3} - 1\right] \qquad \text{(15-25)}$$

and again $W > 0$, consistent with the fact that the work is carried out on the gas by the external agent.

Figure 15-10 shows both the isotherm (dashed curve) and the adiabat (solid curve) through the initial point P_0, V_0. For an isothermal compression the work W carried out by the external agent is numerically equal to the area under the isotherm between the

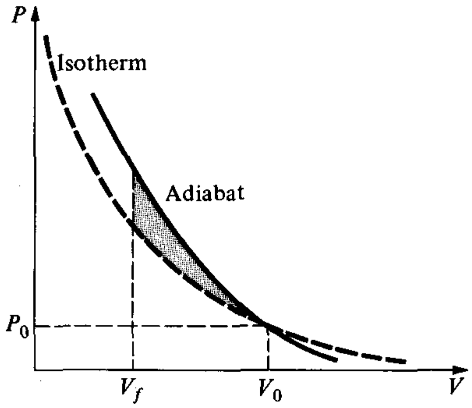


Figure 15-10

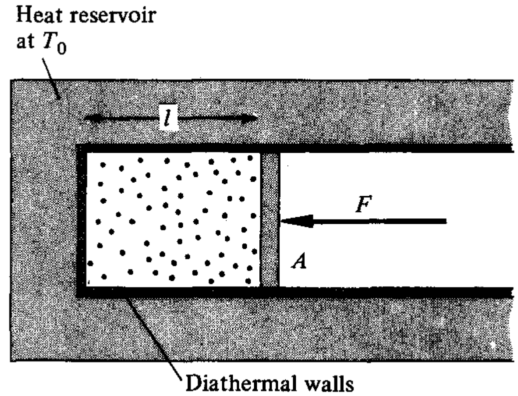


Figure 15-11

two vertical lines at V_0 and V_f (see Example 15-4). Correspondingly, the work carried out during the adiabatic compression is given by the corresponding area under the adiabat. Hence the work required to carry out the adiabatic compression exceeds that required for the isothermal compression by the shaded area in the figure.

Example 15-9 Figure 15-11 shows a dilute gas confined to a cylinder of cross-sectional area A and length l, with a movable piston at one end. Assume that the walls are diathermal and that the system is in thermal contact with a *heat reservoir* at a temperature T_0. That is, assume that the gas is effectively in contact with a system so large that its temperature does not change noticeably even when it gives up or absorbs ordinary amounts of heat. For example, the air in our atmosphere is a heat reservoir under ordinary circumstances. Suppose that suddenly a very large force F with the property $F \gg PA$, where P is the gas pressure, compresses the gas to half its volume.

(a) How much work is carried out on the gas?

(b) Describe in qualitative terms what happens to this energy.

Solution

(a) The work W carried out by the external agent is

$$W = \frac{Fl}{2}$$

since the force F pushes the piston a distance $l/2$.

(b) Immediately after the gas has suffered this sudden compression it is not in thermal equilibrium. Gradually, as a consequence of intermolecular collisions and collisions of the molecules with the confining walls, thermal equilibrium at the reservoir temperature T_0 will be reestablished (see Section 13-3). According to (15-14), the energy of a dilute gas varies only with its temperature T. Hence, since the gas is dilute, its energy after the compression is the same as it was originally, namely $\frac{3}{2}NkT_0$. This implies that the work $Fl/2$ carried out on the gas goes completely into increasing the energy of the reservoir. None of it is retained by the gas.

15-10 Coefficient of thermal expansion

Consider, in Figure 15-12a, a thin strip of a material at a temperature T_0 and having a length l_0, and suppose that it is heated or cooled down so that its

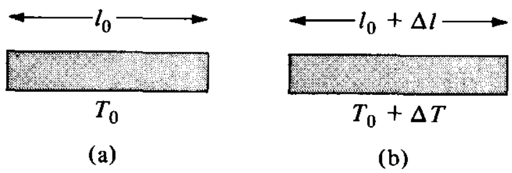


Figure 15-12

temperature becomes $(T_0 + \Delta T)$. As a result of this change in temperature, in general, the length of the strip will change by a small amount; call it Δl. Provided that the variation in temperature ΔT is not too large, we find that this change of length is small and is typically of the order of 0.01 percent.

Experiment shows that for a given material the change in length Δl is directly proportional to both the temperature change ΔT and to the original length l_0. Hence we may write

$$\Delta l = \alpha l_0 \Delta T \tag{15-26}$$

where α is a coefficient of proportionality that varies from substance to substance and is known as the *coefficient of linear expansion.* Table 15-2 lists the coefficients of expansion of a variety of materials at a reference temperature of 273 K. Since, in general, the coefficient of linear expansion of most substances varies with temperature, it is necessary to specify the reference temperature. Qualitatively speaking, the listed values for α are applicable for temperature changes ΔT of the order of 10 K to 50 K. For some materials the coefficient of linear expansion is sensitively dependent on the degree of purity of that substance, and for these cases the values for α in the table represent an average over several samples.

Table 15-2 Values of the coefficient of linear expansion

Substance	α (per degree Celsius)
Aluminum	2.2×10^{-5}
Bronze	1.8×10^{-5}
Copper	1.7×10^{-5}
Gold	1.4×10^{-5}
Nickel	1.3×10^{-5}
Silver	1.7×10^{-5}
Stainless steel	1.5×10^{-5}
Tin	2.3×10^{-5}
Tungsten	0.42×10^{-5}

To extend these ideas to the case of a three-dimensional solid or a liquid, consider in Figure 15-13a a rectangularly shaped isotropic and homogeneous solid at a temperature T_0. Let l_1, l_2, and l_3 represent its linear dimensions. If the temperature of the solid is raised by a small amount ΔT, then according to

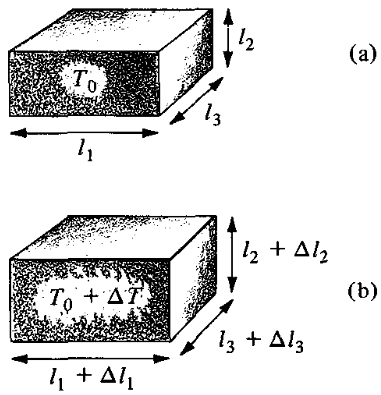


Figure 15-13

(15-26) its dimensions will change by the respective amounts Δl_1, Δl_2, and Δl_3, as follows:

$$\begin{aligned} \Delta l_1 &= \alpha l_1 \, \Delta T \\ \Delta l_2 &= \alpha l_2 \, \Delta T \\ \Delta l_3 &= \alpha l_3 \, \Delta T \end{aligned} \tag{15-27}$$

where α is the coefficient of *linear expansion* of the material and Δl_1, Δl_2, and Δl_3 are as defined in Figure 15-13b.

Let us calculate the change in volume ΔV of the solid as a result of this change in temperature. Assuming that the changes in length are very small compared to the original lengths, we may write

$$\begin{aligned} \Delta V &= (l_1 + \Delta l_1)(l_2 + \Delta l_2)(l_3 + \Delta l_3) - l_1 l_2 l_3 \\ &= l_1 l_2 \Delta l_3 + l_2 l_3 \Delta l_1 + l_1 l_3 \Delta l_2 \end{aligned}$$

where the quadratic and cubic change-of-length terms have been dropped. Substituting into this formula for Δl_1, Δl_2, Δl_3 by use of (15-27), we obtain

$$\Delta V = 3\alpha V_0 \Delta T \tag{15-28}$$

where $V_0 \equiv l_1 l_2 l_3$ is the original volume. Note that the fractional change in volume $\Delta V / V_0$ of a homogeneous isotropic solid is directly proportional to the temperature change ΔT. This time, however, the coefficient of proportionality is three times the coefficient of linear expansion.

A formula analogous to (15-28) is also applicable for liquids. Experiment shows that the change in volume ΔV of a liquid of volume V when its temperature is raised from some reference temperature T_0 to a temperature $(T_0 + \Delta T)$ is

$$\Delta V = \beta V \, \Delta T \tag{15-29}$$

The proportionality constant β is known as the *coefficient of volume expansion* and generally varies from liquid to liquid. Typically, β is 100 times greater than α.

Experiments show that, in general, both the coefficient of linear expansion

α in (15-26) and the coefficient of volume expansion β in (15-29) are *positive.* Physically this means that substances tend to expand when they are heated. However, there is one very striking exception to this rule. This exception is ordinary water in the temperature range near 4°C. As the temperature of water is raised above 4°C its volume increases in the same way as that of most substances. However, if it is cooled *below* 4°C, then, unlike most substances, the volume of a given amount of water *increases.* This means that for temperatures below 4°C the coefficient of volume expansion for water must be negative. In Figure 15-14a the variation of $\Delta V/V$ for water as a function of temperature is shown. Note that for temperatures higher than 4°C the curve has a positive slope, but that below this temperature the slope is negative. This is to be contrasted with the situation for a normal liquid, as shown in Figure 15-14b, where the corresponding curve is a straight line with a positive slope numerically equal to the coefficient of volume expansion β for that liquid.

At temperatures below 0°C, water freezes to form ice. In this process it expands. It is left as an exercise to confirm the fact that as a consequence of the anomalous behavior of the coefficient of volume expansion for water as depicted in Figure 15-14a, the water in a lake freezes from the top down. On the other hand, if the coefficient of volume expansion of water varied as in Figure 15-14b, so that water behaved as a normal liquid, then the lake would freeze starting from the bottom up.

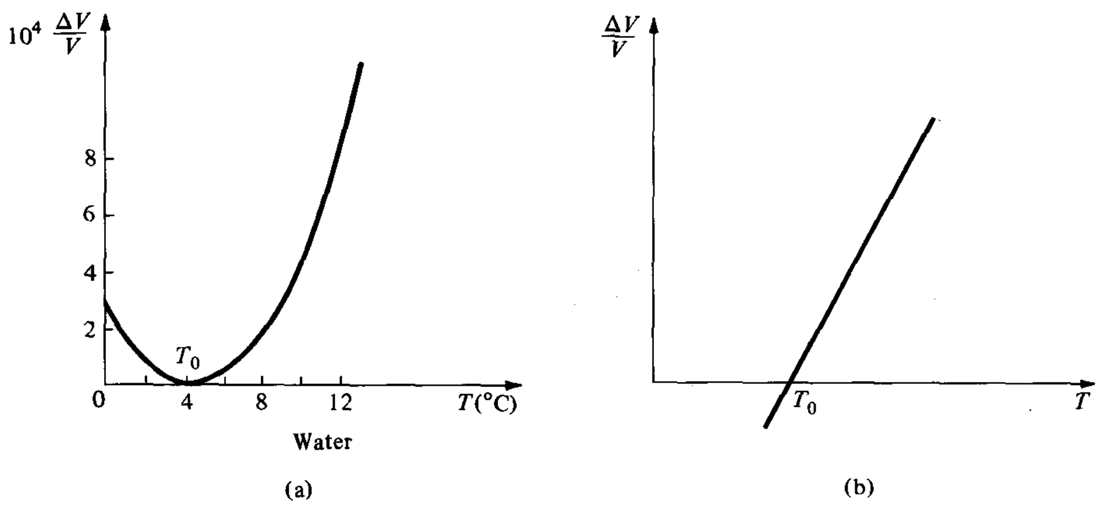


Figure 15-14

Example 15-10 A thin rod of bronze has at 0°C a length of precisely 1 meter. How much longer is it when it has a temperature of 50°C?

Solution In terms of (15-26) and Table 15-2, the parameter values are $\alpha = 1.8 \times 10^{-5}/°\text{C}$, $l_0 = 1.0$ meter, $\Delta T = 50°\text{C}$, since a temperature interval of 50 K is the same as 50°C. Hence

$$\Delta l = \alpha l_0 \, \Delta T = (1.8 \times 10^{-5}/°\text{C}) \times (1 \text{ m}) \times (50°\text{C}) = 9.0 \times 10^{-4} \text{ m} = 0.9 \text{ mm}$$

or, in other words, the rod's length is increased by almost a millimeter.

Example 15-11 At a temperature of 20°C it is found that a certain bridge has a length of 150 meters. Assuming that it is composed mainly of steel, calculate the change in length of this bridge when the temperature has fallen to 0°C.

Solution In terms of (15-26) and Table 15-2, $l_0 = 150$ meters, $\alpha = 1.5 \times 10^{-5}/°\mathrm{C}$, and $\Delta T = -20°\mathrm{C}$. Thus

$$\begin{aligned}\Delta l = \alpha l_0 \, \Delta T &= (1.5 \times 10^{-5}/°\mathrm{C}) \times (150 \text{ m}) \times (-20°\mathrm{C}) \\ &= -4.5 \times 10^{-2} \text{ m}\end{aligned}$$

The minus sign reflects the fact that the bridge has actually contracted as a result of this temperature drop.

Example 15-12 At a temperature of 20°C a thin, straight strip of aluminum is rigidly connected to a similarly shaped strip of tungsten, as in Figure 15-15a. What does this bimetallic structure look like at 0°C and at 40°C?

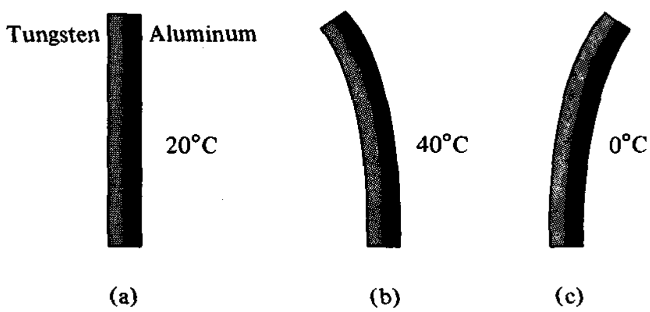


Figure 15-15

Solution According to Table 15-2, the value of α for aluminum is greater than is the corresponding value for tungsten. Accordingly, since l_0 at 20°C is the same for both, it follows from (15-26) that at 40°C the length of the aluminum strip will be greater than that of the tungsten. Hence at the temperature of 40°C the structure will assume a bent shape as in Figure 15-15b, with the center of curvature on the tungsten side. Correspondingly, at the temperature of 0°C the sign in (15-26) for Δl will be negative and this time the tungsten strip will be larger than the aluminum one. Consequently, the structure must bend, as in Figure 15-15c, with the center of curvature on the aluminum side.

15-11 Summary of important formulas

The equation of state of a gas in thermal equilibrium is an equation relating the pressure, the volume, and the temperature of the gas. At sufficiently low densities, the equations of state of *all* gases satisfy the ideal gas law

$$P = \frac{NkT}{V} = nkT \tag{15-10}$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant. This formula can also be written

$$PV = \mu RT \qquad (15\text{-}11)$$

where μ represents the number of moles of the gas and where R, the *gas constant,* is the product of Boltzmann's constant and Avogadro's number. It has the value $R = 8.31$ J/mole-K.

The thermal velocity v_{th} of the molecules of a dilute gas is related to its temperature T by

$$\frac{1}{2}mv_{th}^2 = \frac{3}{2}kT \qquad (15\text{-}13)$$

where m is the mass of a molecule. For a monatomic gas we may express this equivalently as

$$E = \frac{3}{2}NkT \qquad (15\text{-}14)$$

where E is the total energy, which for an ideal gas is the same as its kinetic energy, and where N is the total number of molecules in the gas.

If a thin rod has a length l_0 when its temperature is T_0, then when its temperature is changed by the amount ΔT the associated change in its length Δl is

$$\Delta l = \alpha l_0 \Delta T \qquad (15\text{-}26)$$

where the proportionality constant α is called the coefficient of *linear expansion.*

QUESTIONS

1. Define or describe briefly what is meant by the following: **(a)** thermometric property; **(b)** isothermal expansion; **(c)** adiabatic compression; **(d)** diathermal wall; **(e)** monatomic; **(f)** equation of state; **(g)** thermalization; and **(h)** heat reservoir.
2. What is the zeroth law of thermodynamics? Describe the role played by this law in defining a temperature scale.
3. Suppose that two distinct physical systems are in thermal equilibrium with each other. Must their energy be the same? Explain.
4. Consider the apparatus in Figure 15-3. **(a)** When the gas is heated, what happens to the mercury levels in the three arms A, B, and C if the flexible tube is not moved? **(b)** What happens to the mercury level in these three arms if under the same circumstances the gas is cooled down?
5. Describe what happens to the mercury level in the three arms A, B, and C in Figure 15-3 if the flexible tube is raised. Why must the mercury levels in arms A and B always be at the same level regardless of how the flexible tube is moved?
6. Describe how you could use the apparatus in Figure 15-3 as a *con-tant-pressure* thermometer; that is, one where the correlation is made between the volume of the gas and its

"degree of hotness" at fixed pressure. What would be the definition analogous to (15-4) for such a thermometer?

7. Explain in microscopic terms why the temperature throughout a dilute gas in thermal equilibrium must be the same. What would happen if the gas in one half of the container were hotter than that in the other?
8. Why must the slope of an isotherm—that is, a plot at fixed temperature of the pressure against volume—of a gas be negative? To what would a positive slope correspond physically? Would this violate a law of physics?
9. What is the physical mechanism underlying the fact that the air in a basketball or in an automobile tire becomes hot when in normal usage?
10. Explain in physical terms why the pressure of the gas in an automobile tire increases after the automobile has been driven for a while.
11. In a certain process suppose that the volume of a dilute gas is doubled. Describe what happens to its pressure and temperature if the process is **(a)** isothermal and **(b)** adiabatic.
12. A rigid vessel containing a dilute gas is shaken vigorously. Assuming that the walls of the vessel are insulated, is the energy of the gas changed? What about its temperature?
13. What is the thermal velocity of the molecules of a dilute gas at absolute zero? What can you say about the velocity of the individual molecules under this circumstance? Do you think this physical situation is ever realized?
14. Explain in physical terms why the temperature at which water boils decreases with increasing elevation, so that, for example, at sea level boiling water feels hotter to the touch than it does in the mountains.
15. Write down what the equation of state for an ideal gas would be if the temperature in this formula were expressed in terms of the Celsius scale. What would it look like if the temperature were measured on the absolute Fahrenheit scale?
16. According to Table 15-1, the temperature at the surface of the sun is 6000 K. Can you think of an operational way for justifying this statement?
17. Making use of the data in Table 15-1, determine whether or not it is possible to show *experimentally* that ^{4}He becomes a superfluid at 2.18 K by making use of a constant-volume gas thermometer with ^{3}He as a working substance.
18. What changes, if any, would have to be made in the values of the coefficients of linear expansion in Table 15-2 if instead of using for ΔT the Celsius scale we had used: **(a)** The Kelvin scale? **(b)** The absolute Fahrenheit scale? **(c)** The Fahrenheit scale?
19. Since most substances expand on being heated, what corrections, if any, have to be made in order to use the apparatus in Figure 15-3 as a constant-volume thermometer?
20. Suppose that a thin copper rod, when in equilibrium with ice, has a length of 100.0 cm and, when in equilibrium with boiling water, it has a length of 100.2 cm. Design a thermometer and an associated temperature scale by use of this rod.
21. Why is it that the water in a lake freezes first at the top? Present an argument to show that if the coefficient of volume expansion for water had the "normal" form in Figure 15-14b, then ice would form first at the *bottom* of a lake. What would happen to this ice subsequent to its formation?

PROBLEMS

1. The body temperature of a healthy human being is 98.6°F. Express this temperature on the following scales: (a) absolute Fahrenheit; (b) Kelvin; and (c) Celsius.
2. For what temperature do the readings on the Fahrenheit and the Celsius scales have the same numerical value? Do you think there is any significance to this temperature?
3. Calculate the temperature at which the readings on the following pairs of scales have the same numerical value: (a) Kelvin-absolute Fahrenheit; (b) Kelvin-Fahrenheit; and (c) Celsius-absolute Fahrenheit.
4. A dilute gas at a temperature of 0°C contains 2 moles. (a) What is the temperature of the gas on the Kelvin scale? (b) How many molecules are in the gas? (c) What is the energy of the gas assuming it is monatomic?
5. Argon, at a pressure of 1 atm and a temperature of 10°C, occupies a volume of 30 liters.
   (a) How many moles of gas are there?
   (b) What is the particle density of the gas?
   (c) What is the energy of the argon?
6. Consider again the argon in Problem 5.
   (a) If its volume is halved isothermally, what is its final pressure?
   (b) If it were heated isobarically so that its temperature rose to 40°C, what would be its final volume?
   (c) If it were heated isochorically (at constant volume) to a temperature of 50°C, what would be its final pressure?
7. Consider 1.5 moles of ^{4}He at a temperature of 300 K and occupying a volume of 25 liters. Calculate its energy:
   (a) Initially.
   (b) Finally, if its volume is halved isothermally.
   (c) Finally, if it is heated isobarically so that its temperature rises to 330 K.
   (d) Finally, if it is heated isochorically (at constant volume) to a temperature of 400 K.
8. A bubble of oxygen occupies a volume of 50 cm^3 after it escapes from a submarine that lies at a depth of 75 meters below the surface of the ocean.
   (a) What is the pressure of the oxygen initially?
   (b) Assuming that the temperature stays the same, what is the volume of this bubble just as it reaches the surface?
9. If P, T, and E are, respectively, the pressure, the temperature, and the energy of an ideal gas, show that if the gas undergoes an adiabatic compression, then at each stage of the compression the following quantities are constant: (a) $PT^{-5/2}$ and (b) $PE^{-5/2}$. Assume the gas to be monatomic.
10. (a) On a graph of V against T draw several isobars of 1 mole of an ideal gas. What happens to a given isobar as the gas pressure is raised?
    (b) On the same graph, draw the adiabats; that is, the lines $VT^{3/2} =$ constant.
11. Repeat both parts of Problem 10, but this time draw the isochores (fixed volume) and the adiabats in a P-T diagram for the ideal gas.

*12. The first-order correction to the ideal-gas law, (15-10), for moderately high densities is

$$P = nkT[1 + nB(T)]$$

where $B(T)$ is known as the *second virial coefficient*. Studies in statistical mechanics show that $B(T)$ may

be expressed in terms of an intermolecular potential $V(r)$ by

$$B(T) = 2\pi \int_0^\infty r^2\,dr \times \left[1 - \exp\left\{\frac{-V(r)}{kT}\right\}\right]$$

(a) Calculate $B(T)$ for the special choice:

$$V(r) = \begin{cases} V_0;\ r \leqslant a \\ 0;\ r \geqslant a \end{cases}$$

(b) Evaluate $B(T)$ in the limit $V_0 \gg kT$ and show that if $na^3 \ll 1$, then the equation of state for the ideal gas, (15-10), is correct.

13. One mole of a monatomic dilute gas is allowed to expand isothermally until its volume is doubled. If its initial temperature is 20°C and its initial pressure is 0.8 atm, calculate **(a)** its initial volume; **(b)** its final temperature; and **(c)** its final pressure.

14. Repeat all three parts of Problem 13, but assume this time that the expansion takes place adiabatically.

15. An automobile tire has a volume of $4.0 \times 10^4\ \text{cm}^3$ and contains air at a pressure of 2 atm and at a temperature of 50°F.

(a) What is the particle density of the air in the tire?

(b) After the automobile has been driven for a while, suppose that the temperature of the air has risen to 100°F. What is now the pressure of the air in the tire, assuming that its volume change is negligible?

(c) By how much has the energy of the air in the tire been increased as a result of this heating process in (b)? Assume the gas to be monatomic.

16. Consider a sample of ordinary air at a temperature of 30°C.

(a) What is the thermal velocity of the ^{4}He atoms in this sample?

(b) What is the thermal velocity of the oxygen molecules in this sample?

(c) What is the average kinetic energy of the nitrogen molecules?

17. A half mole of argon at 0°C is admitted to a 20-liter vessel containing originally 0.75 mole of ^{4}He at a temperature of 30°C.

(a) What is the initial pressure of the ^{4}He?

(b) What is the final temperature of the mixture?

(c) What is the final pressure of the mixture and the partial pressure due to each constituent?

18. For the situation shown in Figure 15-7, suppose that $N_A = N_B = 2.5 \times 10^{23}$, $T_A = 280$ K, and $T_B = 330$ K, that both gases are ^{4}He, and that each initially occupies a volume of 20 liters.

(a) What are the initial pressures, P_A and P_B, of the two gases?

(b) What is the final temperature when thermal equilibrium has been established?

(c) Calculate the final pressure of each gas.

19. Suppose, in Figure 15-7a, that the gases on both sides of the partition are ^{4}He and are characterized by the parameter values $N_A = N_B = 3 \times 10^{23}$, $T_A = 260$ K, $T_B = 300$ K, and $V_A = V_B/4 = 25$ liters. Assume now, however, that the diathermal partition between the gases is *freely movable.*

(a) Calculate the initial pressures of the two gases.

(b) Explain why it is that when thermal equilibrium is established the densities of the two gases will be the same. Calculate this value. (*Hint*: Why will the pressures of the two gases be the same at equilibrium?)

(c) What are the final temperature and pressure of the two gases?

20. Consider the same situation as in Problem 19, but suppose this time that the movable piston is *not* diathermal but rather is an insulator.
   (a) Show by use of the conservation-of-energy law that the sum of the final two temperatures after thermal equilibrium is established is the same as the sum of the initial two temperatures.
   (b) Show that the ratio of the final volumes is the same as that of the final temperatures.

21. One mole of dilute krypton is initially at a temperature of 0°C and is compressed adiabatically so that its volume is decreased by 30 percent. (a) What is the final temperature of the gas? (b) What is the change in the energy of the krypton in this process?

22. Consider a room full of air that has a volume of 350 cubic meters and is at a temperature of 20°C and a pressure of 1 atm.
   (a) How much energy is required to increase the temperature of the air to 30°C? Assume that air is monatomic.
   (b) To what volume must the air be compressed adiabatically to achieve a temperature rise of 10°C?
   (c) What is the final pressure of the air in (a)? What is it in (b)?

23. Consider a space capsule that has an available volume of 50 liters and is originally filled with 0.8 mole of helium at 20°C.
   (a) What is the original pressure of the ^{4}He?
   (b) After having orbited in space for a while, the capsule gradually cools off until it achieves a final temperature of − 200°C. What is the pressure now? How much energy has the ^{4}He lost in this cooling process?

24. A thin rod of copper has a length of 40 cm. What is the change in its length if its temperature is raised by 100 K? Use the value for the linear-expansion coefficient in Table 15-2.

25. Using the values of α in Table 15-2, calculate the difference in length between the strips of aluminum and tungsten in Figures 15-15b and 15-15c, assuming that at 20°C each strip is precisely 20 cm long.

26. Show that the period P of a simple pendulum is increased by the amount $\frac{1}{2}\alpha P\,\Delta T$ when its temperature is raised by ΔT. The symbol α here represents the linear-expansion coefficient for the supporting string. (*Hint*: For a simple pendulum of length l, $P = 2\pi\sqrt{l/g}$.)

27. Show, by using arguments similar to those used to derive (15-28), that the change ΔA in the area A of a material in the form of a planar, rectangular figure when its temperature is raised by ΔT is $\Delta A = 2\alpha A\,\Delta T$, where α is the coefficient of linear expansion of that material.

28. A very thin, spherical shell has a radius of 15 cm and is composed of aluminum. If its temperature is raised by 20 K, calculate (a) the change in its surface area; (b) the change in its volume; and (c) the change in its radius.

29. Repeat Problem 28, but assume this time that the sphere is solid.

30. Show, by starting with (15-29), that the change in density $\Delta\rho$ of a liquid of density ρ when its temperature is changed by ΔT is $\Delta\rho = -\beta\rho\,\Delta T$. Discuss the reason for the minus sign.

31. (a) The moment of inertia I of a solid sphere of radius b and mass m about a diameter is $\frac{2}{5}mb^2$. Assuming that the material of the sphere has the linear-expansion coefficient α, show that if the temperature is raised by ΔT, then the change of its moment of inertia is $2\alpha I\,\Delta T$. (b) Show that if a sphere is rotating with

angular velocity ω about an axis through its center, then if its temperature is raised by ΔT, its angular velocity will change by $-2\alpha\omega\,\Delta T$.

32. Starting with the definition for the moment of inertia I about a given axis in (12-4), show that if the temperature of a rigid body is raised by ΔT, then the change in I is $2\alpha I\,\Delta T$.
33. The temperature of a "pure" silver dollar is increased by 100 K. Calculate the relative change, in percent, of its **(a)** area; **(b)** thickness; **(c)** volume; and **(d)** density.
34. According to Archimedes' principle, if a body of volume V_0 is submerged in a fluid of mass density ρ_0, then it is buoyed up by the force $\rho_0 V_0 g$. Show that if the temperature of the body and the fluid are raised by ΔT, then the buoyant force at the new temperature is

$$\rho_0 V_0 g[1+(3\alpha-\beta)\Delta T]$$

where α is the linear-expansion coefficient of the body and β is the coefficient of volume expansion of the fluid. (*Hint*: Use the result of Problem 30 and (15-29).) Since, in general, $\beta \gg \alpha$, does this correspond to an increase or decrease in the buoyant force?

*35. A liquid with a coefficient of volume expansion β occupies, as shown in Figure 15-16, a hollow, spherical shell of radius r_0, composed of a material with a coefficient of linear expansion α. A small capillary tube of cross-sectional area A_0 is attached to an opening in the sphere.

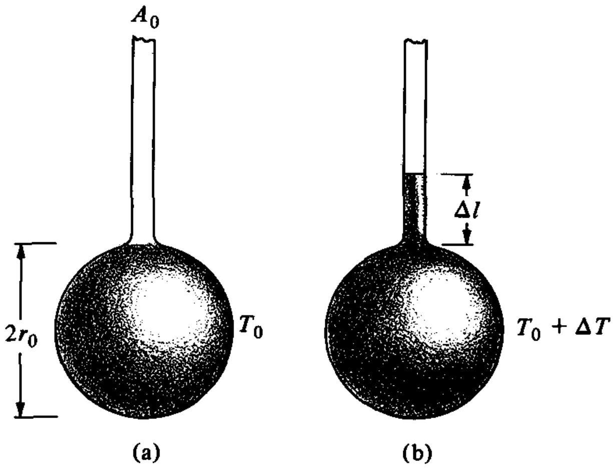


Figure 15-16

Suppose that the temperature of the system is raised from an original temperature T_0 to $(T_0+\Delta T)$.

(a) What is the change in volume of the spherical shell?

(b) What is the change in volume of the liquid?

(c) Show that the height Δl of the liquid in the capillary column (see Figure 15-16b) is given by

$$\Delta l = \frac{4\pi r_0^{\,3}}{3A_0}(\beta-3\alpha)\,\Delta T$$

(*Note*: For liquids at room temperature, β has in typical cases the values 1.8×10^{-4}/K for mercury, 1.1×10^{-3}/K for ethyl alcohol, while α for solids is typically of the order of 10^{-5}/K. Thus in general $\beta \gg 3\alpha$ or, in other words, the change in volume of the container in Figure 15-16 is negligible compared with the corresponding volume change of the liquid.)

16 Heat and the first law of thermodynamics

Thermodynamics is the only physical theory of universal content concerning which I am convinced that within the framework of the applicability of its concepts it will never be overthrown.

ALBERT EINSTEIN (1879–1955)

Die Energie der Welt ist konstant;
Die Entropie der Welt strebt einem Maximum zu.

RUDOLPH CLAUSIUS (1822–1888)

16-1 Introduction

In Chapter 15 the word "heat" was introduced to denote that entity which passes between two systems, initially at different temperatures, after they are placed in thermal contact. In order to be able to give a more quantitative definition of this concept, it is necessary to introduce the *first law of thermodynamics.* This law and the associated *second law of thermodynamics,* which we shall study in Chapter 17, comprise the theoretical basis that underlies much of our understanding of the behavior of thermodynamic systems.

In simplest terms, the first law is a statement of the conservation of energy for thermodynamic systems and the fact that heat is a form of energy. To

view it merely as the statement of energy conservation, however, is a gross oversimplification. The first law is very much more than that, and it plays an extremely important role not only in physical science but in many areas of the chemical and the biological sciences as well.

16-2 State functions

Consider a fluid composed of N identical molecules. To characterize a state of this system, it is necessary, in general, to specify $6N$ parameters; for example, the initial positions and velocities of all the molecules. By contrast, if the system is in thermodynamic equilibrium, then a state of this system can be uniquely specified in terms only of its temperature T and volume V. For example, if the two gases in Figure 16-1 are characterized by the same values for the thermodynamic variables of V and T and each has the same number N of molecules, then they are in precisely the same thermodynamic state even though from a microscopic point of view the positions and momenta of their constituent molecules may differ vastly in detail. Similarly, two vessels, each containing the same amount of water and each in thermal equilibrium with the air in a room, are also in the same thermodynamic state despite the possibly different motions of their constituent molecules. In each of these cases, since the two systems are in thermal equilibrium and are characterized by identical values for N, V, and T, they are in precisely the same thermodynamic state.

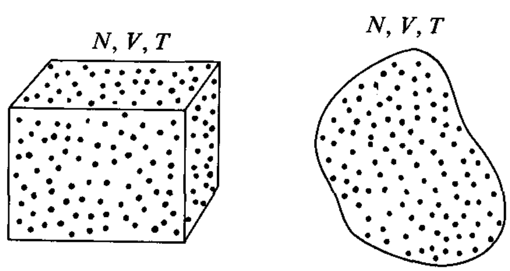


Figure 16-1

Let us now make these notions more precise. A liquid or a gas is said to be in a state of *thermal* or *thermodynamic equilibrium* provided that it can be uniquely characterized macroscopically by its volume V, its temperature T, and the number N, and the chemical nature of its constituent atoms and molecules. Any fluid that cannot be described in this simple way in terms only of the three variables N, V, and T is said not to be in thermal equilibrium.

Consider now a fluid in an equilibrium state characterized by T, V, and N. A physical quantity which depends only on the state of this system—that is, only on T, V, and N—is said to be a *state function*, provided that its value is determined uniquely only by the values for T, V, and N for that state. In

particular, the value of a state function cannot depend on the process by means of which the system attained the given state in the first place. If, for example, the volume and temperature of a gas change as a result of its undergoing some process, then the associated change in the value of *any state function* depends only on the initial and final values for T and V and not on the details of the process used to bring about this change. By contrast, if the change in a function of the state of a gas, as it undergoes a process, depends also on the details of this process, then that function is *not* a state function.

An important example of a state function is the total energy of a gas. According to (15-14) the total energy E of a monatomic ideal gas is

$$E = \frac{3}{2} NkT \tag{16-1}$$

Since this depends only on N and T (for a dense gas it would also depend on V) and has nothing to do whatsoever with the nature of the process used to bring the gas into this state, it follows that E is a state function. For the same reason, the volume and the temperature of a gas are also state functions.

As a special case, consider the N molecules of an ideal monatomic gas initially in a state characterized by point A on the T_1-isotherm in Figure 16-2. According to (16-1), the energy E_A of the gas in this state is $3NkT_1/2$. If the gas is now taken to the state corresponding to point B on the T_2-isotherm, then its energy E_B will be $3NkT_2/2$. Hence, the change in energy ΔE of the gas is the difference

$$\Delta E = E_B - E_A = \frac{3}{2} Nk(T_2 - T_1) \tag{16-2}$$

and this is the same regardless of which process (out of the infinite number of possible ones) is used to bring about this change. For example, the gas could be taken along the path ACB by first being compressed isobarically to a volume V_B and then having its pressure increased isochorically until it is on the T_2-isotherm. Or it could be taken along the path ADB by first raising its temperature isochorically to T_2 and then compressing it isothermally to

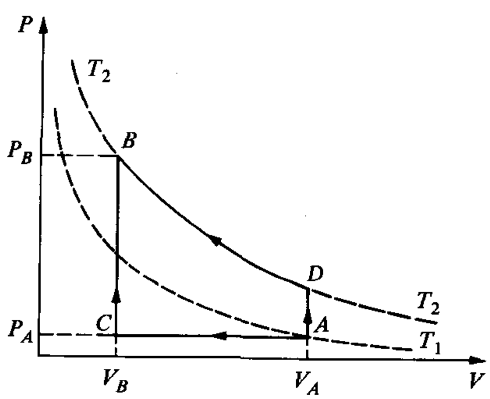


Figure 16-2

the final volume V_B. Or the gas could be taken from A to B by use of any one of a number of *nonequilibrium* processes—for example, by suddenly compressing it to a volume V_B and then allowing it to come to thermal equilibrium at the final temperature T_2. The point is that, regardless of which of these or any other processes that are used to take the gas from A to B, since E is a state function the change in energy of the gas is the same for all of them.

By contrast to this, the work carried out on the gas as it goes from A to B is *not* a state function. For even though W depends on the values of the temperature and volume of the gas initially and finally, it *also* depends on the nature of the process used to bring about this change. Hence, *W is not a state function.* For example, if the process corresponds to the path ADB in Figure 16-2, then W is the area under the T_2-isotherm between the vertical lines at V_A and V_B. On the other hand, if the process corresponds to the path ACB, then W is the area of the rectangle below the horizontal line AC and bounded by the same two vertical lines. Since these two areas are different, so is the work required in the two cases. Hence the work W required to take the gas from A to B cannot be represented in terms of a state function.

16-3 The first law of thermodynamics

Consider, in Figure 16-3a, a thermodynamic system at some initial energy E and suppose that a certain amount of work W is carried out on this system so that, as shown, its energy becomes $(E + \Delta E)$. For any given macroscopic process, this work W can be calculated, in principle, by applying the ideas of Newtonian mechanics. If, for example, the system is a gas at pressure P and it is compressed so that its volume changes by a small amount ΔV, then this work is $(-P\,\Delta V)$. Or if the system is a glass of water and the process consists of stirring it, then W is the product of the distance the stirrer moves and the component of the required force along the direction of motion. Unfortunately, it is not possible to calculate the associated energy change ΔE of the system so directly.

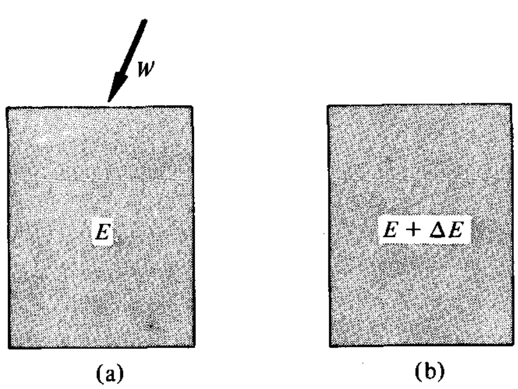


Figure 16-3

As a first guess, one might argue that for the situation in Figure 16-3 the energy change ΔE is the same as is the work W carried out on the system. Indeed, this is precisely what we found in Section 15-9 for the adiabatic compression of an ideal gas. However, except for very special circumstances of this type, the energy change of a system undergoing some process is *not*, in general, equal to the work carried out on it during this process.

We define the *heat* Q added to a system when it undergoes some process to be the difference between the change in its energy ΔE and the work W carried out on it. According to basic ideas of mechanics, the energy of an isolated system is always conserved. Hence Q can be thought of as that part of the change of the energy of a system which is *not* produced by the work W carried out on it. In mathematical terms, then,

$$\Delta E = W + Q \tag{16-3}$$

where we have adopted the sign convention $Q > 0$ to mean heat is *added to* the system from some external source. Correspondingly, a negative value for Q means that during the process of interest the system gives off heat.

Now as noted above, the work W carried out on a system undergoing some process can be calculated in terms of purely macroscopic quantities. However, even with (16-3) available it may still not be clear how the change in energy ΔE of a system and the heat Q added to it can be calculated independently. It might seem, for example, that (16-3) is merely a relation between the two macroscopic quantities ΔE and Q, neither one of which is itself separately measurable. The fact that this is not at all the case, and this is a very important point, has to do with the fact that the energy E of a thermodynamic system is a state function. It is for this reason that (16-3) is the powerful predictive tool known as the first law of thermodynamics and not simply the definition of heat.

In order to see how the energy difference ΔE between any two states of a system can be determined by use of (16-3) consider, for example, the two processes shown schematically in Figure 16-4. The initial state here involves some ice in a glass of water at 0°C and the final state is the original water and the melted ice at 0°C. Two ways for bringing about this change of state are shown in the figure. One of these processes involves heating the ice until it is melted and the second consists of surrounding the container with insulating walls and stirring the icewater mixture vigorously until the ice is all melted. Since the energy E of the system is a state function, the energy change ΔE is the same for both processes. Therefore, assuming that the walls of the container are insulating and thus prevent the leakage of heat into or out of the system, it follows that the work W involved in stirring must be the same as the energy difference ΔE. With this energy difference ΔE thus established once and for all, (16-3) can then be used to calculate the heat Q added to this system for any other process connecting these two states. For since ΔE is known and since W can be calculated for any given process by use of mechanics, it follows that the heat Q added to the system is determined uniquely by (16-3).

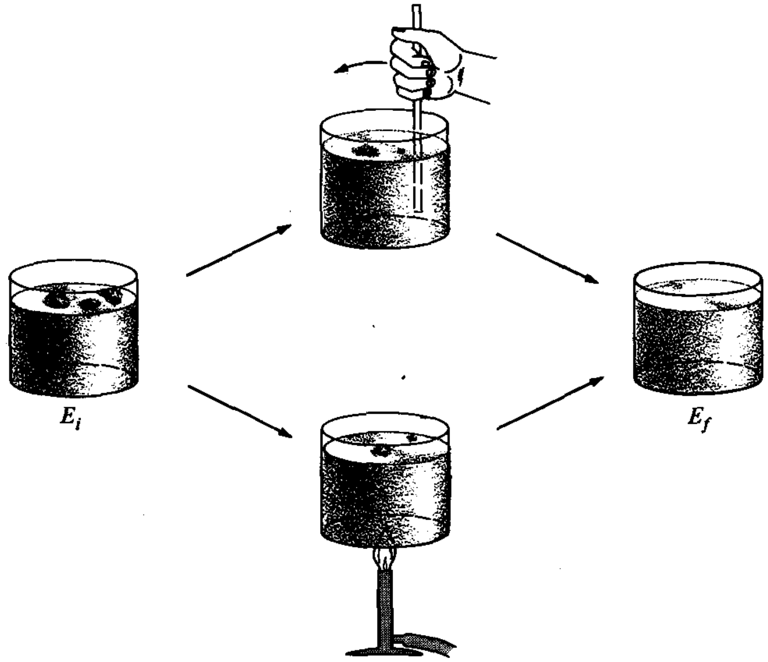


Figure 16-4

And it is in this way that (16-3) can be used to determine both ΔE and Q independently.

To summarize, then, the *first law of thermodynamics* states:

If a thermodynamic system is taken from an equilibrium state i of energy E_i to an equilibrium state f of energy E_f by any process, then the work W carried out on the system in this process and the heat Q added to it are related by

$$E_f - E_i = W + Q \tag{16-4}$$

with $(E_f - E_i)$ the same for all processes connecting the two states.

Note that it is not necessary for the system to be in thermal equilibrium at each intermediate state of this process. The only requirement that (16-4) be applicable is that the system be in thermodynamic equilibrium in its initial and final states.

16-4 Applications of the first law

To obtain further insight into the significance and usefulness of the first law, in this section we apply it to a number of elementary physical situations.

Example 16-1 In an experiment it is found that it takes 333 joules of mechanical work in stirring to melt 1 gram of ice at 0°C to produce water at the same temperature.

Assuming that the system is thermally insulated, calculate:
(a) The change in energy ΔE of the ice.
(b) The amount of heat required to achieve the same transformation.
(c) The amount of heat that must be added to the ice if the same transformation of ice takes place, but only 200 joules of mechanical work is carried out.

Solution
(a) In terms of (16-3), we are given the data $Q = 0$ and $W = +333$ joules. Hence

$$\Delta E = W + Q = 333\text{ J} + 0 = 333\text{ J}$$

(b) Since $\Delta E = 333$ joules for this change of state and since no mechanical work is carried out on the system, a second application of (16-3) yields

$$Q = \Delta E - W = 333\text{ J} - 0 = 333\text{ J}$$

(c) This time the parameter values are $\Delta E = 333$ joules and $W = 200$ joules. Hence

$$Q = \Delta E - W = 333\text{ J} - 200\text{ J} = 133\text{ J}$$

The fact that this is positive implies that energy in the form of heat must be added to the system from external sources.

Example 16-2 In Figure 16-5 suppose that 0.5 mole of an ideal monatomic gas is kept at atmospheric pressure $P_0 = 1.0 \times 10^5\text{ N/m}^2$ in a cylinder with a freely movable piston and that a certain amount of heat Q is added to the gas in such a way that at all times it remains in thermal equilibrium with its pressure fixed at the value P_0. If the volume of the gas increases from the initial value $V_i = 1.0 \times 10^4\text{ cm}^3$ to a final value $V_f = 2.0 \times 10^4\text{ cm}^3$, calculate:
(a) The amount of work W carried out on the gas.
(b) The initial temperature T_i and the final temperature T_f of the gas.
(c) The change in energy ΔE of the gas.
(d) The amount of heat Q added to the gas.

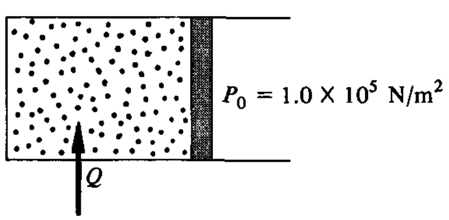


Figure 16-5

Solution
(a) Since the pressure is constant, the process is isobaric. Hence

$$\begin{aligned} W &= -\int_{V_i}^{V_f} P\,dV = -P_0(V_f - V_i) \\ &= -(1.0 \times 10^5\text{ N/m}^2) \times [2.0 \times 10^{-2}\text{ m}^3 - 1.0 \times 10^{-2}\text{ m}^3] \\ &= -1.0 \times 10^3\text{ J} \end{aligned}$$

The minus sign reflects the fact that the gas is itself carrying out positive work.

(b) Making use of the equation of state for the ideal gas we have

$$T_i = \frac{P_0 V_i}{\mu R} = \frac{(1.0 \times 10^5 \text{ N/m}^2) \times 10^{-2} \text{ m}^3}{0.5 \text{ mole} \times 8.31 \text{ J/mole-K}} = 241 \text{ K}$$

Since the process is isobaric, it follows by use of the equation of state for the ideal gas that V and T are directly proportional to each other throughout the process. Since the volume of the gas is doubled, so must be the temperature. Hence

$$T_f = 2 \times 241 \text{ K} = 482 \text{ K}$$

a result that also follows directly from the equation of state.

(c) Making use of (16-2), we find for the change in the energy of the gas

$$\begin{aligned} \Delta E &= \frac{3}{2} Nk(T_f - T_i) = \frac{3}{2} \mu R(T_f - T_i) \\ &= \frac{3}{2} \times (0.5 \text{ mole}) \times (8.31 \text{ J/mole-K}) \times (482 \text{ K} - 241 \text{ K}) \\ &= 1.5 \times 10^3 \text{ J} \end{aligned}$$

(d) Solving (16-3) for Q and inserting the above values for the energy change ΔE and the work W, we find that

$$Q = \Delta E - W = 1.5 \times 10^3 \text{ J} - (-1.0 \times 10^3 \text{ J}) = 2.5 \times 10^3 \text{ J}$$

Example 16-3 Consider an ideal gas at temperature T which is slowly compressed isothermally from an initial volume V_0 to a final volume $V_0/3$. How much heat must flow into the gas from external sources?

Solution For an isothermal compression, the temperature T of the gas does not change. Hence, since the energy E of an ideal gas depends only on T and not on volume (see (16-2)), it follows that $\Delta E = 0$. Substituting the given values $V_i = V_0$; $V_f = V_0/3$ into (15-24), we find that

$$W = \mu RT \ln 3$$

and substituting these values for ΔE and W into (16-3), we obtain

$$Q = \Delta E - W = -\mu RT \ln 3$$

The final minus sign here reflects the fact that heat flows *out* of the gas and into the temperature reservoir whose presence is required to maintain the gas at the fixed temperature T.

Example 16-4 A monatomic ideal gas containing N molecules is compressed very slowly and adiabatically from an initial state characterized by V_0 and T_0 to a final one characterized by V_f and T_f. Show that consistent with our definition of an adiabatic process, the heat Q added to the gas vanishes exactly.

Solution Substituting the given parameters into (15-25), we find that

$$W = \frac{3}{2} NkT_0 \left[\left(\frac{V_0}{V_f} \right)^{2/3} - 1 \right]$$

where we have used the equation of state, $P_0V_0 = NkT_0$, to eliminate P_0. Now, according to (15-21), the product $V^{2/3}T$ is constant throughout an adiabatic compression. Hence

$$\left(\frac{V_0}{V_f}\right)^{2/3} = \frac{T_f}{T_0}$$

so that W may be expressed in the form

$$W = \frac{3}{2}Nk(T_f - T_0)$$

On the other hand, according to (16-2), the change in energy ΔE of the gas in this process is

$$\Delta E = \frac{3}{2}Nk(T_f - T_0)$$

Hence it follows by use of (16-3) that no heat is added to the gas.

Example 16-5 A certain physical system S (for example, a real gas or a liquid) is placed in thermal contact (via a diathermal wall) with μ moles of a monatomic ideal gas at an original temperature T_0 (Figure 16-6). Assuming that the walls surrounding the two systems are rigid and insulating and that the common final temperature is T_f, calculate the change in energy ΔE_S of system S and the heat Q_S added to it.

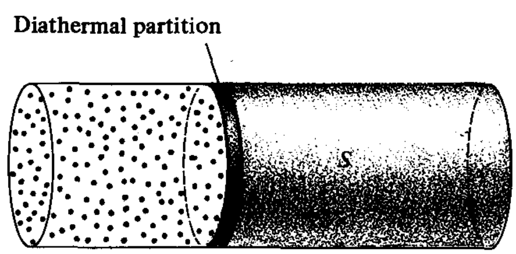


Figure 16-6

Solution Since the two systems are isolated, the sum of their energies cannot change in this process of their coming to equilibrium. Hence

$$\Delta E_S = -\Delta E_g$$

where ΔE_g is the energy change of the ideal gas. Moreover, according to (16-2), $\Delta E_g = 3\mu R(T_f - T_0)/2$, and thus

$$\Delta E_S = -\frac{3}{2}\mu R(T_f - T_0)$$

Since the volume of neither the gas nor of system S changes in this process of thermalization, it follows that no mechanical work is carried out on either one. According to (16-3), then, the heat Q_S added to system S is

$$Q_S = \Delta E_S - W_S = -\frac{3}{2}\mu R(T_f - T_0)$$

since $W_S = 0$.

If the temperature of the ideal gas rises, so that $(T_f - T_0) > 0$, then the heat added to system S is negative according to this formula, and heat flows out of S into the gas. A corresponding conclusion follows if the gas temperature drops during the process of thermalization.

It is interesting to note how by use of the technique exemplified in this example, the energy change ΔE of real gases and liquids can be measured for certain transformations. Thus, a considerable amount of information can be obtained of the energy function for any real fluid by allowing it to come into thermal equilibrium with an ideal gas and making use of the known form of the energy function for the latter.

16-5 Units of heat

Experiments involving heat and its transport between physical systems at different temperatures were carried out long before it was clearly understood that heat is a form of energy. Were it not for this circumstance there would be no need to belabor the point that the units of heat are the same as those of energy. However, because of this historical fact, it is necessary to define two other units of heat which are in common usage today. These are the *calorie* (cal) and the *British thermal unit* (Btu).

Originally, the calorie was defined to be the quantity of heat required to raise the temperature of 1 gram of water from 14.5°C to 15.5°C. Correspondingly, the Btu was originally defined to be the amount of heat required to raise the temperature of 1 lb of water (at sea level) from 63°F to 64°F. Neglecting the slight difference in temperature between these two definitions, in the problems it is shown that these units are related by

$$1 \text{ cal} = 3.968 \times 10^{-3} \text{ Btu} \tag{16-5}$$

The kilocalorie (kcal), which is 10^3 cal, is also frequently utilized. The "calorie" used to measure the energy content of food is actually the kilocalorie.

A relation of considerable importance in thermodynamics is the *mechanical equivalent of heat*. This relates the units of the calorie and British thermal unit on the one hand and the joule on the other. In principle, we could obtain this relation experimentally by putting water at 14.5°C into an insulating container and determining how much mechanical work must be carried out to raise its temperature to 15.5°C. An experiment of this very sort was first carried out by James Joule (1818–1889), who found that it took approximately 4 joules of mechanical work to produce 1 cal of heat. Subsequent and more precise measurements have refined this estimate, and the value accepted today is

$$1 \text{ cal} = 4.186 \text{ J} \tag{16-6}$$

A related unit is the *thermochemical calorie*; this is *defined* to be precisely 4.1840 joules.

If (16-5) and (16-6) are combined, it becomes possible to express the British thermal unit in mechanical terms. Table 16-1 summarizes this and similar relations. According to the table, for example, 1 Btu is the same as 252.0 cal, as well as 1055 joules and 778.0 ft-lb.

Table 16-1 Units of heat and energy

	J	*cal*	*Btu*	*ft-lb*
1 J	1	0.2389	9.480×10^{-4}	0.7376
1 cal	4.186	1	3.968×10^{-3}	3.087
1 Btu	1055	252.0	1	778.0
1 ft-lb	1.356	0.3239	1.285×10^{-3}	1

Example 16-6 Experiment shows that 80 cal of heat are required to melt 1 gram of ice at 0°C to produce water at the same temperature. How much mechanical work in joules must be supplied to melt 2 grams of ice if only 40 cal of heat is available?

Solution In order to melt 2 grams of ice, 160 cal are required. Since only 40 cal are available in the form of heat, it follows from the first law that 120 cal of mechanical work must be carried out. According to Table 16-1, 1 cal is the same as 4.2 joules, and hence

$$120 \text{ cal} = 120 \text{ cal} \times 4.2 \text{ J/cal} = 5.0 \times 10^2 \text{ J}$$

units of mechanical work are required.

16-6 Heat capacity

Experiment shows that if heat is added to a thermodynamic system its temperature generally rises. Conversely, if heat is extracted from a system a drop in temperature is normally observed. To quantify this correlation between the heat Q added to a thermodynamic system and its associated temperature change ΔT it is convenient to introduce the notion of *heat capacity.*

Consider a thermodynamic system of total mass m at an initial temperature T. Suppose than an infinitesimal amount of heat ΔQ is added to or extracted from it and that, as a result, its temperature changes by a small amount ΔT. Physically, we expect that the more heat is added the larger will be ΔT, and that for a fixed ΔQ, the more massive the system is, the smaller ΔT will be. Bearing these features in mind, we define the *specific heat* c of the given system by

$$\Delta Q = cm \, \Delta T \tag{16-7}$$

where, to be consistent with the experimental fact that the addition of heat is invariably associated with a rise in temperature, c is an inherently positive quantity.

The units of specific heat are those of energy per unit mass per degree. Hence, a typical unit for c is the cal/g-K. Since the calorie is equal to the amount of heat required to raise 1 gram of water by 1 K, it follows that the specific heat of water is

$$c = \frac{\Delta Q}{m\,\Delta T} = \frac{1\ \text{cal}}{1\ \text{g} \times 1\ \text{K}} = 1\ \text{cal/g-K}$$

A quantity related to the specific heat is the *molar heat capacity, C.* Consider a homogeneous system of mass m composed of molecules of molecular weight A. The number μ of moles of the system is m/A, provided that the mass m of the substance is expressed in grams. Assuming this to be the case, (16-7) may be expressed in the equivalent form

$$\Delta Q = C\mu\,\Delta T \tag{16-8}$$

where the molar heat capacity C is defined by

$$C = cA \tag{16-9}$$

For example, since the specific heat of water at 14.5°C is 1 cal/g-K and since the molecular weight of water is 18, it follows by use of (16-9) that its molar heat capacity is 18 cal/mole-K.

The definitions in (16-7) and (16-8) assume that both ΔQ and ΔT are sufficiently small that neither C nor c varies appreciably throughout the temperature interval ΔT. More generally, both the specific heat c and the molar heat capacity C may depend on the temperature of the system. The appropriate generalizations of (16-7) and (16-8) for these cases are

$$Q = \mu \int_{T_1}^{T_2} C(T)\,dT = m \int_{T_1}^{T_2} c(T)\,dT \tag{16-10}$$

where Q is the net heat added to m grams or μ moles of the system while its temperature is changed from T_1 to T_2.

In addition to varying with temperature, the specific heat of a given physical system also depends, in general, on the nature of the process used to add heat. For example, if an ideal gas undergoes an isochoric process involving the addition of a certain amount of heat Q, the rise in its temperature ΔT will in general be different from the corresponding temperature rise if the gas were heated at constant pressure. As a rule, it is necessary to associate a different heat capacity with each distinct process to which a physical system might be subjected. Of particular importance in this connection are C_V, the molar heat capacity at constant volume, and C_P, the molar heat capacity at constant pressure. Associated with each of these are the specific heats at constant volume c_V and at constant pressure c_P, which are related to the corresponding molar heat capacities in accordance with (16-9). In general, the molar heat capacity at constant volume is not equal to that at constant pressure. The notion of a molar heat capacity or of specific heat without further qualification is ambiguous and should never be used.

As will be considered briefly in Problem 24, above a certain temperature, known as the Debye temperature, the molar heat capacity at constant volume C_V of many crystalline solids has the value $3R$, with $R \equiv 8.31$ J/mole-K, the gas constant. For these substances, then, $C_V \cong 6.0$ cal/mole-K. The fact that at high temperatures, $C_V = 3R$, is known as the law of Dulong and Petit.

Example 16-7 How much heat is required to raise the temperature of a block of aluminum of mass 1 kg from 600 K to 700 K? Assume that the aluminum is held at constant volume and that $C_V = 3R$.

Solution Since the atomic weight of Al is 27, the number μ of moles in 1 kg is

$$\mu = \frac{10^3 \text{ g}}{27 \text{ g/mole}} = 37 \text{ moles}$$

Substituting this and the value $C_V = 3R$ into (16-8), we obtain

$$\begin{aligned}\Delta Q &= \mu C_V \, \Delta T = 37 \text{ moles} \times (6.0 \text{ cal/mole-K}) \times 100 \text{ K} \\ &= 2.2 \times 10^4 \text{ cal}\end{aligned}$$

Example 16-8 A 2-mole chunk of lead at a temperature of 700°C is dropped into 200 grams of water contained in a 1-kg silver bowl. Assuming that initially both the water and the bowl are at 0°C, calculate the final temperature of the mixture if the heat exchange takes place at constant volume and there is no heat leakage out of the system. Assume $C_V = 3R$ for both the lead and the silver.

Solution Since the molecular weight of water is 18, it follows that 200 grams correspond to $200/18 = 11$ moles. Similarly, since the molecular weight of silver is 108, 1 kg is the same as $1000/108 = 9.3$ moles. The molar heat capacity of water, as we saw above, is 18 cal/mole-K, and that of Pb and Ag is $3R$.

Now since the lead slug, the silver bowl, and the water comprise an isolated system, and since all volume changes are negligible, it follows that whatever heat is gained or lost by one of these three elements comes at the expense of the other two. Hence

$$\text{heat lost by the lead} = \text{heat gained by the water} + \text{heat gained by the bowl}$$

Now if t is the final equilibrium temperature in degrees Celsius of the system, then

$$\begin{aligned}\text{heat lost by the lead} &= \mu C_V \, \Delta T = 2 \text{ moles} \times (6.0 \text{ cal/mole-K}) \times (700°\text{C} - t) \\ &= 12(700 - t) \text{ cal}\end{aligned}$$

since the size of the degree is the same on the Celsius and the Kelvin scales, and since (16-8) involves only temperature differences. In the same way we find that

$$\begin{aligned}\text{heat gained by the water} &= \mu C_V \, \Delta T \\ &= 11 \text{ moles} \times (18 \text{ cal/mole-K}) \times (t - 0°\text{C}) \\ &= 200t \text{ cal}\end{aligned}$$

as well as

$$\begin{aligned}\text{heat gained by bowl} &= \mu C_V \, \Delta T \\ &= 9.3 \text{ moles} \times (6.0 \text{ cal/mole-K}) \times (t - 0°\text{C}) \\ &= 56t \text{ cal}\end{aligned}$$

Combining these formulas, we obtain

$$12(700 - t) = 200t + 56t$$

and this yields

$$t = 31°\text{C}$$

16-7 Molar heat capacity of the ideal gas

The purpose of this section is to derive the molar heat capacities, C_V and C_P, of the ideal gas.

Let us first calculate C_V for the monatomic gas. Consider, in Figure 16-7, a rigid vessel of volume V containing μ moles of an ideal gas at an original temperature T and pressure P. If a small amount of heat ΔQ is added to the gas, both P and T will increase. Since the volume V of the gas is fixed, the work W carried out on—or performed by—the gas vanishes. It follows from the first law that the energy increase ΔE of the gas is in this case precisely the same as the heat added, ΔQ. Therefore, since the energy E of the gas is $\frac{3}{2}\mu RT$ it follows that the energy change $\Delta E = 3\mu R\ \Delta T/2$. Hence

$$\Delta Q = \Delta E = \frac{3}{2}\mu R\ \Delta T$$

and comparing this with the definition of molar heat capacity in (16-8) we find that

$$C_V = \frac{3}{2}R \tag{16-11}$$

This means that at constant volume it takes $3R/2$ units of heat to raise the temperature of 1 mole of a monatomic ideal gas by 1 K.

The corresponding calculation of the molar heat capacity at constant pressure is somewhat more complex. This time, as shown in Figure 16-8, imagine adding ΔQ units of heat to μ moles of a monatomic ideal gas at an initial temperature T and confined to a vessel with a freely movable piston so that its pressure is fixed at a certain value P. As the infinitesimal heat ΔQ is

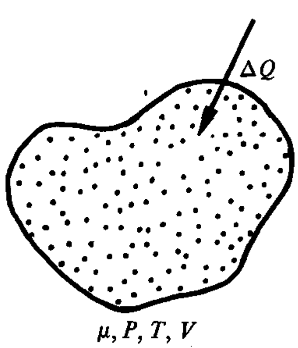


Figure 16-7

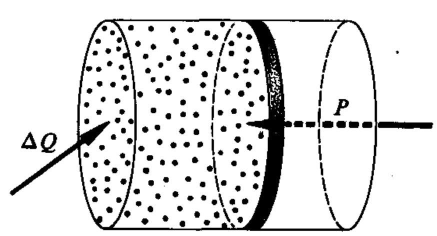


Figure 16-8

added to the gas, in general both its volume and temperature will increase. Let us call these changes ΔV and ΔT, respectively. According to the equation of state, these changes are related by $P\,\Delta V = \mu R\,\Delta T$, and hence the work ΔW carried out on the gas is

$$\Delta W = -P\,\Delta V = -\mu R\,\Delta T$$

As for the constant volume case, the energy change ΔE is $\Delta E = \frac{3}{2}\mu R\,\Delta T$ and substituting these formulas for ΔE and ΔW into (16-3), we find that

$$\Delta Q = \Delta E - \Delta W = \frac{3}{2}\mu R\,\Delta T - (-\mu R\,\Delta T) = \frac{5}{2}\mu R\,\Delta T$$

Comparison with the definition in (16-8) then leads to

$$C_P = \frac{5}{2}R \tag{16-12}$$

so that it takes 5/3 as much heat to raise the temperature of a monatomic ideal gas by 1 K at constant pressure than it does at constant volume.

Table 16-2 lists experimental values for C_P and for C_V for a variety of dilute gases, including monatomic as well as diatomic and polyatomic gases. In general, for monatomic gases the predicted values for the molar heat capacities in (16-11) and (16-12) agree very well with experiment. On the other hand, for diatomic and polyatomic gases, (16-11) and (16-12) are in disagreement with experiment. Note that nevertheless the difference in the specific heats $(C_P - C_V)$ which according to (16-11) and (16-12) has the constant value R seems to be satisfied by these more complex gases as well. The final column of the table lists the values for the ratio C_P/C_V of the molar heat capacities for the various gases. It is customary to use the symbol γ for this ratio. Reference to the table shows that for monatomic gases γ has the value $1.67 = 5/3$, as predicted, while for diatomic gases it appears to have the value $1.40 = 7/5$. For polyatomic gases it seems to behave in a more complex way. It is shown in the problems that if the formula for the energy in (16-1) is replaced by

$$E = \frac{5}{2}\mu RT \qquad \text{(diatomic gas)} \tag{16-13}$$

Table 16-2 Molar heat capacities of dilute gases

Type	*Gas*	C_P/R	C_V/R	$(C_P - C_V)/R$	$\gamma = C_P/C_V$
Monatomic	He	2.50	1.50	1.00	1.67
	A	2.50	1.50	1.00	1.67
Diatomic	H_2	3.46	2.45	1.01	1.41
	N_2	3.49	2.50	0.99	1.40
	O_2	3.53	2.54	0.99	1.40
Polyatomic	CO_2	4.44	3.42	1.02	1.30
	SO_2	4.85	3.78	1.07	1.29

then the experimental values for diatomic molecules in Table 16-2 result. The physical explanation is that if allowance is made for the fact that in addition to the energy of its center of mass a diatomic molecule can also rotate about two mutually perpendicular axes (perpendicular to the axis of the molecule), then (16-13) is a natural consequence. In this way it is possible to account for the values for C_P and C_V for dilute diatomic as well as monatomic gases.

Example 16-9 How much heat is required to increase the temperature of 1 mole of dilute helium by 100 K if:

(a) The volume of the gas is kept fixed?

(b) Its pressure is kept constant?

Solution

(a) Making use of (16-8), and noting that helium is a monatomic gas and thus (16-11) is applicable, we find that

$$\begin{aligned}\Delta Q &= \mu C_V\,\Delta T = 1\text{ mole} \times (3R/2) \times 100\text{ K}\\ &= 1.5\text{ moles} \times (8.31\text{ J/mole-K}) \times 100\text{ K}\\ &= 1.25 \times 10^3\text{ J} = 300\text{ cal}\end{aligned}$$

(b) Since the ratio of the heats added in these two cases is the same as that of the corresponding ratio of molar heat capacities, it follows that at constant pressure the required heat is

$$\Delta Q = \left(\frac{C_P}{C_V}\right) \times 300\text{ cal} = 500\text{ cal}$$

16-8 Heat conduction

Consider, in Figure 16-9, a homogeneous slab of rectangular cross-sectional area A and of length l. Suppose that the left-hand face is in thermal contact with a heat reservoir (not shown) at a temperature T_2 and that the right-hand face is similarly in contact with the heat reservoir at a lower temperature T_1. For simplicity assume that the remaining four sides of the block are covered by an insulating material so that there is no leakage of heat out of the sides. Physically, we expect that heat will flow along the slab from the hot end at

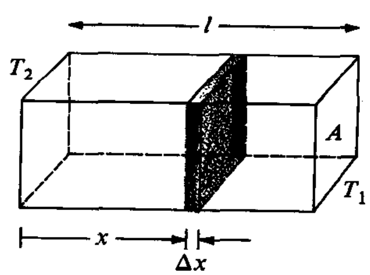


Figure 16-9

temperature T_2 to the cooler one at T_1. We say that this flow is *steady-state*, provided that no heat accumulates at any point inside the slab.

Let ΔQ represent the amount of heat that under steady-state conditions is transported along the slab from the hotter to the cooler reservoir in an infinitesimal time interval Δt. Experiment shows that ΔQ varies directly with the length of the time interval Δt, the cross-sectional area A, and the temperature difference $(T_2 - T_1)$, and inversely with the length l. Introducing the symbol k for the coefficient of proportionality, we may summarize these results by the formula

$$\Delta Q = \frac{k\, \Delta t\, A(T_2 - T_1)}{l}$$

or, equivalently, as

$$\frac{\Delta Q}{\Delta t} = kA\frac{(T_2 - T_1)}{l} \tag{16-14}$$

The proportionality constant k is called the *coefficient of heat conduction*, or the *thermal conductivity*, and its value depends, in general, on the material used. The unit of k is energy per unit time per unit length per degree. Table 16-3 lists representative values for a number of metals at 18°C.

Table 16-3 Thermal conductivities at 18°C

Substance	Al	Cu	Au	Fe	Pb	Hg	Sn	Zn
k (cal/cm-s-K)	0.48	0.92	0.70	0.16	0.08	0.02	0.16	0.27

One interesting application of (16-14) is to the problem of the temperature distribution along a slab, such as that in Figure 16-9. Under steady-state conditions, there is no accumulation of heat anywhere in the slab, so the rate of heat flow $\Delta Q/\Delta t$ must be the same everywhere along the direction of heat flow. In particular, then, applying (16-14) to the thin slab of material at a distance x from the left face of the slab and having a thickness Δx, we find that

$$\frac{\Delta Q}{\Delta t} = kA\frac{T(x) - T(x + \Delta x)}{\Delta x} \tag{16-15}$$

where $T(x)$ is the temperature at this distance x. In the limit as the thickness Δx of the slab becomes very small, this relation becomes

$$\frac{\Delta Q}{\Delta t} = -kA\frac{dT}{dx} \tag{16-16}$$

Note the minus sign! Its presence is required by the fact that heat flows from regions of high temperature to lower ones. The quantity dT/dx is known as the *temperature gradient*.

One of the advantages of having the heat-flow equation available in the form (16-16) is that it can be used to calculate the variation in temperature

along the slab. If we assume steady-state conditions, each of the three quantities $\Delta Q/\Delta t$, k, and A are constant along the slab, and (16-16) may be integrated to

$$T(x) = -\left(\frac{1}{kA}\frac{\Delta Q}{\Delta t}\right)x + \alpha$$

with α an integration constant. Recalling that at $x = 0$, $T(0) = T_2$, we find that $\alpha = T_2$ and hence

$$T(x) = T_2 - \left(\frac{1}{kA}\frac{\Delta Q}{\Delta t}\right)x$$
$$= T_2 - \frac{1}{l}(T_2 - T_1)x$$

where in the second equality we have substituted for $\Delta Q/\Delta t$ by use of (16-14). Figure 16-10 shows a plot of this linear temperature distribution along the slab.

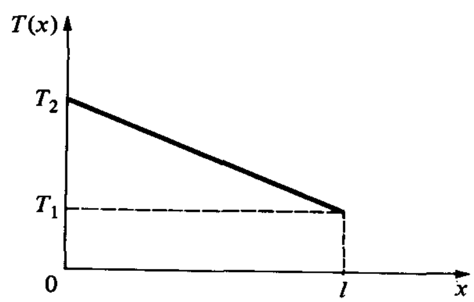


Figure 16-10

Example 16-10 Calculate the rate at which heat is transported along an aluminum block of length 20 cm and of cross-sectional area 4 cm^2 if a temperature difference of 50 K is maintained between opposite faces.

Solution In terms of the notation of (16-14), we are given the following data: $A = 4\text{ cm}^2$, $l = 20\text{ cm}$, and $(T_2 - T_1) = 50\text{ K}$. Hence

$$\frac{\Delta Q}{\Delta t} = \frac{kA}{l}(T_2 - T_1) = \left(0.48\frac{\text{cal}}{\text{cm-s-K}}\right) \times (4\text{ cm}^2) \times \frac{50\text{ K}}{20\text{ cm}}$$
$$= 4.8\text{ cal/s}$$

where the value for k has been taken from Table 16-3.

Example 16-11 Consider a hollow sphere composed of a material of thermal conductivity k and of inner radius R_1 and of outer radius R_2 and suppose that the inner and outer surfaces are maintained at the respective temperatures T_1 and T_2. Calculate the rate $\Delta Q/\Delta t$ at which heat flows to the inner surface.

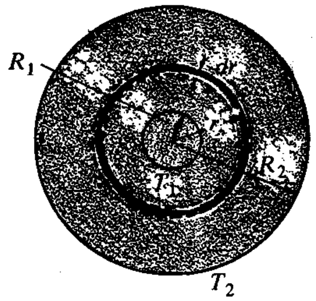


Figure 16-11

Solution Consider, in Figure 16-11, a spherical shell of radius r and of thickness dr. If $T(r)$ is the temperature of this shell it follows from (16-16) that

$$\frac{dT}{dr} = -\frac{1}{kA}\frac{\Delta Q}{\Delta t} = -\frac{1}{4\pi k r^2}\frac{\Delta Q}{\Delta t}$$

since the surface area of the shell is $4\pi r^2$. Integrating and noting that k and $\Delta Q/\Delta t$ are constants, we obtain

$$T(r) = T_2 + \frac{1}{4\pi k}\left(\frac{1}{r} - \frac{1}{R_2}\right)\frac{\Delta Q}{\Delta t}$$

where $T_2 = T(R_2)$ is the appropriate constant of integration. Finally, substituting the value $T(R_1) = T_1$, we obtain the desired result

$$\frac{\Delta Q}{\Delta t} = 4\pi k(T_2 - T_1)\frac{R_1 R_2}{R_1 - R_2}$$

†16-9 The critical point of a fluid

Most substances can exist in three or more distinct physical states, known as *phases*. The most well known of these are the *solid phase*, the *liquid phase*, and the *gaseous phase*. At atmospheric pressure, for example, for temperatures above 100°C, water exists only in its gaseous phase, steam; below 0°C, only in its solid phase, ice; and in between only as ordinary water. For an environment characterized by an appropriate pressure and temperature, corresponding results are true for essentially all known substances. Generally speaking, as the temperature of a substance, originally in its gaseous state, is decreased, it will first liquefy and ultimately become a solid.

By contrast to the hyperbolic isotherms, $PV = \text{constant}$, of the ideal gas, Figure 16-12 shows the isotherms of a real fluid for temperatures T near a particular value T_c known as the *critical temperature*. Experiment shows that, for temperatures $T_1(> T_c)$, if a gas is compressed isothermally its pressure changes in a smooth, though generally nonhyperbolic, way—as represented, for example, by the upper curve in the figure. A similar curve is obtained if an isothermal compression is carried out at the critical temperature T_c itself.

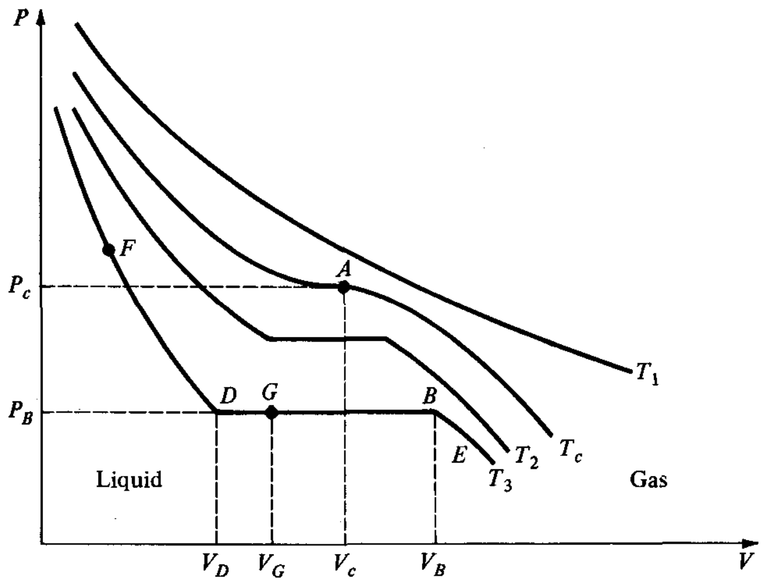


Figure 16-12

However, along this *critical isotherm*, as it is known, there is a certain point A which is characterized by a volume V_c and a pressure P_c—known as the *critical volume* and *critical pressure*, respectively—where the isotherm has simultaneously a point of zero slope and of inflection. That is, at point A,

$$\frac{\partial P}{\partial V} = 0 \qquad \frac{\partial^2 P}{\partial V^2} = 0 \tag{16-17}$$

where the partial derivatives are to be evaluated on the critical isotherm at the *critical point* $V = V_c$ and $P = P_c$. Table 16-4 lists values of the critical-point parameters T_c, P_c, and the critical molar density $\rho_c = N_0 m / V_c$ for some common substances. According to this table, at its critical temperature of 374°C the isotherm for water vapor has a horizontal slope and a point of inflection at a pressure of 218 atm and a molar density of 0.4 g/cm^3. Experiment shows that for temperatures below the critical temperature T_c, most gases will condense to a liquid if sufficient pressure is applied.

Table 16-4 Critical-point parameters

Substance	T_c (°C)	P_c (atm)	$\rho_c = N_0 m / V_c$ (g/cm^3)
H_2O	374	218	0.40
H_2	−240	12.8	0.031
He	−268	2.26	0.069
N_2	−147	33.5	0.31
O_2	−118	49.7	0.43
Air	−141	37.2	0.35
Cl_2	144	76.1	0.57
CO_2	31	73	0.46

Imagine a dilute gas initially in a state corresponding to point E on the T_3-isotherm in Figure 16-12, with $T_3 < T_c$. Suppose the gas is compressed isothermally—by extracting heat as necessary—so that its pressure P rises to the value at point B. At this point, there is a discontinuity in the slope of the isotherm so that to compress the gas isothermally beyond this point it suffices to extract heat without increasing the pressure. In other words, further compression of the gas is now possible at the constant pressure P_B, and the effect of decreasing the gas volume now means that the gas condenses to a liquid. Hence, while the fluid is along the horizontal portion BD of the isotherm it is a mixture of liquid and gas. Finally, when the system is in the state characterized by point D, the gas has condensed completely and is now entirely a liquid. At this point D there is a second discontinuity in the slope of the isotherm. If the liquid is compressed beyond this point, the isotherm rises, but now more steeply than before. This is a reflection of the fact that, in general, liquids are less easily compressed than are gases.

In a similar way, the T_3-isotherm can be retraced backward along the route $FDGBE$. This time, along the horizontal part of the isotherm DB, the liquid *boils* as it goes from the liquid phase to the gaseous phase.

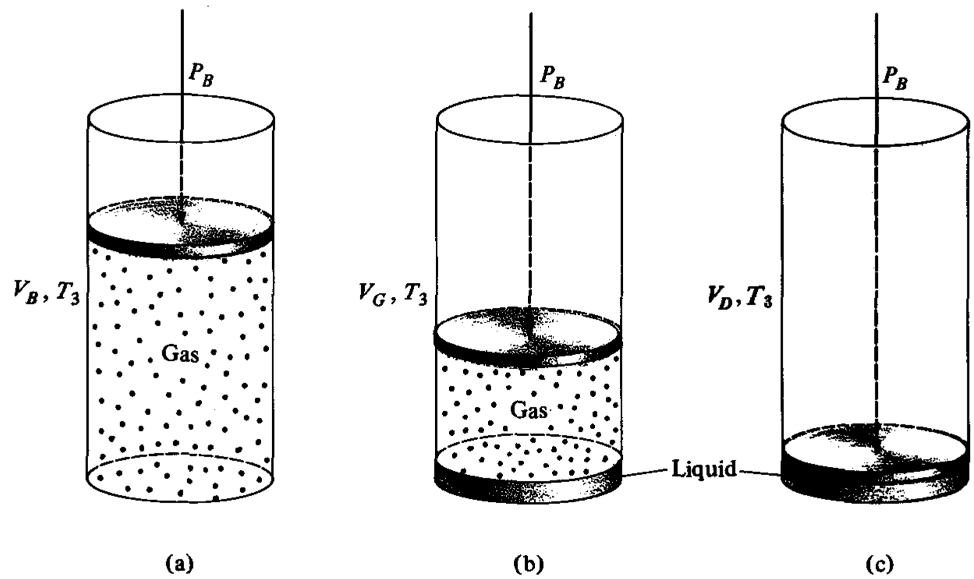


Figure 16-13

Figure 16-13 shows what happens to a fluid along the horizontal portion BGD of the T_3-isotherm. Initally, the system is a gas and is characterized by a volume V_B, a pressure P_B, and a temperature T_3. After some heat has been extracted, the gas will be in a state corresponding to the point G, in which some of the gas has condensed to form a liquid, which because of the influence of gravity settles to the bottom of the container. Finally, as shown in Figure 16-13c, with the extraction of still more heat the system reaches point D, which corresponds to complete condensation to the liquid phase.

†16-10 Heats of fusion, vaporization, and sublimation

The *heat of vaporization*, l_v of a substance, is defined to be the amount of heat required to transform 1 gram of the substance from the liquid to the gaseous state at a fixed temperature and pressure. The temperature, where at a given pressure the liquid vaporizes, is known as its *boiling point*. In terms of Figure 16-12, this means that l_v is the amount of heat required to take 1 gram of a liquid from point D to point B. In general, l_v depends on both the temperature and pressure at which the liquid vaporizes. Physically, l_v represents the energy (per unit mass) that must be supplied to the liquid to overcome the attractive forces that the molecules exert on each other at close range.

Table 16-5 lists the boiling points and the heats of vaporization for various substances at a pressure of 1 atm. According to this table, for example, it takes 51 cal of heat to boil 1 gram of liquid oxygen at atmospheric pressure and at a temperature of $-183°C$.

Table 16-5 Heats of a vaporization and fusion at 1 atm

Substance	*Boiling point* (°C)	*Heat of vaporization* (cal/g)	*Melting point* (°C)	*Heat of fusion* (cal/g)
H_2O	100	540	0.0	80
O_2	-183	51	-219	3.3
N_2	-196	48	-210	6.2
Cu	2600	1760	1080	50
Hg	357	70	-39	2.7
Ag	2200	550	960	24

The same table also lists the melting points of these substances at a pressure of 1 atm and a quantity called the *heat of fusion*. By analogy to the heat of vaporization, the heat of fusion l_f of a substance is defined to be the amount of heat required to cause 1 gram of that substance to melt. In physical terms then l_f represents the energy per unit mass that must be supplied to destroy the energetically more favorable crystalline structure of the solid. At a given pressure, the temperature at which a solid melts is known as its *melting point*. According to Table 16-5, for example, at a pressure of 1 atm and at a temperature of $-39°C$ mercury can be either in a solid or in a liquid state. If it is in a solid phase, then 2.7 cal/g of heat are required to cause it to melt to the liquid state at the same pressure and temperature. Conversely, if it is in the liquid state at this temperature and pressure, then 2.7 cal/g must be extracted to cause it to solidify.

Under restricted circumstances, the addition of heat will also cause certain solids to vaporize directly into the gaseous state. This process is known as *sublimation*. For example, at room temperature and at a pressure of 1 atm,

solid carbon dioxide (also known as "dry ice") sublimates. The *heat of sublimation* l_s is defined to be the heat required to cause 1 gram of a substance to go directly from the solid to the gaseous state.

A *phase diagram* for a substance is a plot of the pressures and temperatures at which it changes its state. Figure 16-14 represents schematically the phase diagram for a substance that can exist in a liquid, a solid, or a gaseous phase.

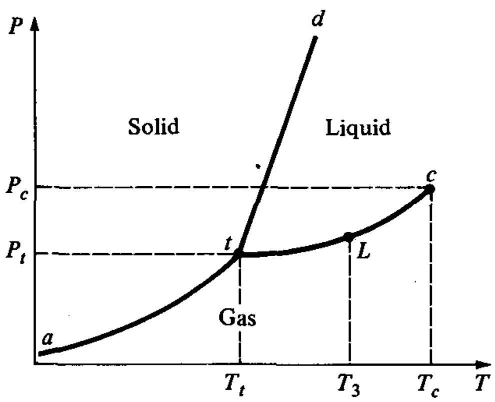


Figure 16-14

For any given pressure P and temperature T the graph shows the phase of the substance. Along the line *ta*, the solid and the gas phase of the substance are in thermal equilibrium with each other, and this line is known as the *sublimation line*. The line *tc* is know as the *vaporization line*, since for values of P and T along *tc* the liquid and the gas phase of the substance are in thermal equilibrium. For example, all points along the horizontal part of the T_3-isotherm in Figure 16-12 are represented by the single point L in the figure. Finally, the line *td* is known as the *melting curve* or the *fusion line*, since for values of P and T along here the solid phase and the liquid phase of the substance are in equilibrium with each other. The endpoint of the vaporization line—that is, the point c in the figure—is the *critical point*. For all temperatures above T_c there is no way of distinguishing between the liquid and the gaseous phase of the substance. The point t at which the sublimation, the fusion, and vaporization lines meet is known as the *triple point*. (Recall in this connection that in defining the constant-volume gas thermometer use was made of the triple point of water, which is associated with the parameter values $T_t = 273.16$ K and $P_t = 0.458$ cm Hg.) The triple-point parameters P_t and T_t, as well as the critical point parameters P_c and T_c, determine the phase diagram of a substance to an appreciable extent. For example, the fact that solid CO_2 sublimates at 1 atm can be seen directly from the phase diagram for this substance by use of the experimental values $P_t = 5.0$ atm and $T_t = -56.6°C$. Hence, a pressure of 1 atm must correspond to a point somewhere along the line *ta* in Figure 16-14, and thus at 1 atm solid CO_2 must sublimate when heat is added.

As the range of pressures and temperatures at which experiments are carried out has been extended, frequently new phases of a substance have

been discovered. In this sense the phase diagram in Figure 16-14 is an oversimplification. For example, Figure 16-15 shows the phase diagram for ^{4}He. Note that there is no sublimation line for ^{4}He and further that there are *two phases* for the liquid state. These are known as HeI and HeII, and they are separated by the so-called λ*-line.* For temperatures $T < T_\lambda = 2.18$ K, gaseous ^{4}He will liquefy to form HeII, while for temperatures above T_λ and below $T_c = 5.2$ K it liquefies to form the HeI phase. HeII has a number of interesting properties, including the ability to flow through tubes of very small diameter ($\leqslant 0.1$ mm) without resistance. It is called a *superfluid.*

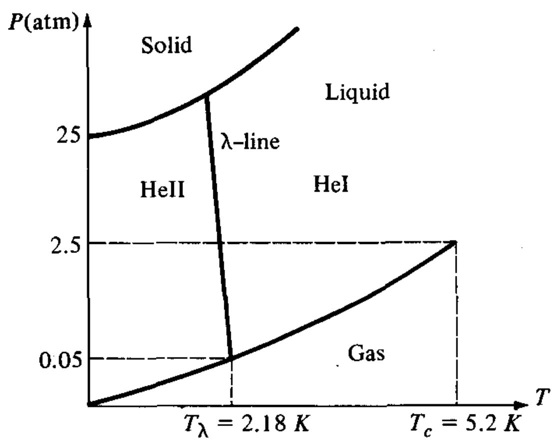


Figure 16-15

16-11 Summary of important formulas

The main result of this chapter is the first law of thermodynamics. If a system is taken from initial equilibrium state i of energy E_i to a final equilibrium state f characterized by the energy E_f, then the energy change $\Delta E = E_f - E_i$ is related to the work W carried out on the system and the heat Q added to it by

$$\Delta E = W + Q \tag{16-3}$$

If an infinitesimal amount of heat ΔQ is added to a system of mass m, then its specific heat c is defined by

$$\Delta Q = cm\,\Delta T \tag{16-7}$$

where ΔT is the associated temperature change. The corresponding formula for the molar heat capacity, C, is

$$\Delta Q = C\mu\Delta T \tag{16-8}$$

where μ is the number of moles in the system. In general, both the molar heat capacity and the specific heat will depend on the nature of the process involved and the symbols C_V and C_P represent the molar heat capacities at constant volume and pressure, respectively.

Suppose in Figure 16-9 the end faces of a body are kept at different temperatures, T_2 and T_1. Then the amount of heat ΔQ transported from one end to the other time in a interval Δt is

$$\frac{\Delta Q}{\Delta t} = kA\frac{(T_2 - T_1)}{l} \tag{16-14}$$

where A is the cross-sectional area of the body and l is its length. The coefficient of proportionality k is called the coefficient of *heat conduction*, or the *thermal conductivity*.

QUESTIONS

1. Define or describe briefly the following: (a) state function; (b) diatomic gas; (c) specific heat; (d) heat; (e) sublimation; and (f) thermal conductivity.
2. Consider a gas consisting of N molecules. Suppose that the gas is initially characterized by the volume and temperature V_1 and T_1, respectively, and that these assume the final values V_2 and T_2 as a result of its being subjected to a certain process. Which of the following quantities depend only on the initial and final states of the gas and which depend on the nature of the process involved: (a) Its volume? (b) Its temperature? (c) The work carried out on it? (d) Its energy? (e) Its pressure? Do not assume that the gas is necessarily ideal.
3. A monatomic gas consists of N molecules. Assuming that it is in thermodynamic equilibrium, explain why it is not necessary to specify $6N$ initial conditions to define a unique macroscopic state. How many conditions are required for this purpose?
4. Explain why only the volume and not the shape of the confining vessel of a gas affects its thermodynamic behavior. Do you think the shape might be important for gas consisting of only, say, 10^6 molecules?
5. Describe the central role played by adiabatic processes in making the first law a useful predictive tool. Illustrate by reference to a specific system.
6. Consider a physical system that is in a certain state such that *all* other states of the system can be reached from this one by an adiabatic process. Describe in operational terms how you would measure the energy differences between all states of this system by carrying out mechanical work.
7. A monatomic ideal gas is taken from a state characterized by V_1 and T_1, to a state characterized by V_2 and T_2. If $V_1T_1^{3/2} = V_2T_2^{3/2}$, describe in detail an adiabatic process connecting these two states.
8. Repeat Question 7, but suppose this time that $V_1T_1^{3/2}$ is not equal to $V_2T_2^{3/2}$.
9. Suppose that a physical system is taken between two thermodynamic states by a process for which the system is *not in equilibrium* at some intermediate state. Does the first law apply to this system under these circumstances? Explain.
10. It is desired to apply the first law to a system which is taken from a state A to a state B but it is found that there is no adiabatic process connecting state A to state B. However, there is an adiabatic process connecting B to A. Is it possible to measure the energy

difference associated with the transition $A \rightarrow B$ by carrying out mechanical work?

11. Would it be correct to define C_V as the amount of heat required to raise the temperature of 1 mole of a substance by 1 K at constant volume? Would it be correct to define C_P in a similar way but at constant pressure?
12. One mole of a substance has its temperature raised by the amount ΔT at constant volume. Would its energy change ΔE be given by $C_V \Delta T$? Would its energy change be $C_P \Delta T$ if the process were carried out at constant pressure?
13. Can either C_P or C_V be negative? What would a negative value mean in physical terms?
14. Explain in physical terms why you would expect C_P to be greater than or equal to C_V at any fixed temperature for any given substance.
15. In Example 16-8, would it not have been more appropriate to have assumed that the process takes place at constant pressure? Do you think the results would have been modified appreciably if this had been done?
16. Using the fact that at 1.0 atm it takes 540 cal/g to convert H_2O at 100°C to steam at the same temperature, estimate the average potential energy between two water molecules. (*Hint*: Assume that steam is sufficiently dilute to be considered an ideal gas.)
17. If a substance is forced to go from B to D along the T_3-isotherm in Figure 16-12, why must heat be extracted? Where does this energy come from?
18. Would there be horizontal parts to the isotherms of a fluid if there were no attractive part in the intermolecular potential? Explain.
19. Reference to Figure 16-14 shows that for temperatures above T_c there seems to be no line of demarcation between the gaseous and the liquid phases of a substance. What would happen to a liquid at a pressure $P > P_c$ and originally at a temperature $T < T_c$ which is heated isobarically to a temperature larger than T_c? Would it become a gas or remain a liquid?
20. Explain, in physical terms, what would happen to a fluid along part of a P-V isotherm that has a positive slope.
21. Explain in terms of intermolecular forces why the heats of fusion in Table 16-5 are generally much smaller than are the heats of vaporization for any given substance.
22. The two opposite ends of a copper block are kept at a fixed temperature difference so that heat flows along the block. If steady-state conditions prevail, is the copper in thermodynamic equilibrium? Is the first law applicable?

PROBLEMS

1. One mole of argon occupies a volume of 20 liters and is at a temperature of − 10°C. Assume that ideal gas conditions prevail.
   (a) What is the pressure of the gas?
   (b) If the gas is allowed to expand isothermally to a volume of 25 liters, how much work must be carried out on it?
   (c) How much heat must be added to the argon in (b) so that the process is isothermal?
2. One gram of gaseous helium is in a container with a movable piston that serves to keep the gas at the fixed pressure of 0.5 atm. Assume that initially the gas temperature is 0°C and that ideal gas conditions prevail.
   (a) How much heat must be added

to raise the temperature slowly to 100°C?
(b) How much work is carried out in this process?
(c) Calculate the initial and final volumes of the gas.

3. A reversible cyclic process is one in which a system is forced to return to its initial state after going continuously through a series of equilibrium states. For example, when a gas undergoes a cyclic process, the final values for V and T are precisely the same as the initial values. Show that in a cyclic process the work carried out *by* the gas is equal to the heat added to it.
4. A space capsule has an available volume of 140 liters and contains 6 moles of air at an original temperature of 0°C. On entering the atmosphere, suppose that the temperature of the air inside the capsule rises to 100°C. Assuming that air is monatomic: (a) How much heat has been added to the air? (b) What is the energy change of the air?
5. What is the change in energy of a system that carries out 50 joules of work and absorbs 50 cal of heat in a certain process?
6. A 3-kg block of aluminum, originally traveling at a velocity of 5 m/s, slides along a rough, horizontal surface until it comes to rest. (a) How much energy is dissipated as heat in this process? (b) If the block absorbs 80 percent of this heat, calculate its temperature rise, assuming that $C_V = 3R$.
7. A bullet of mass 5 grams, and traveling horizontally at a velocity of 100 m/s strikes the bob of a pendulum of mass 2 kg and becomes embedded in it.
   (a) What is the velocity and energy of the bob after the collision?
   (b) How much energy, in calories, is dissipated during this process?
   (c) Assuming that C_V for the bullet and bob is $3R$, calculate the tempeťature rise of the system due to this energy dissipation in (b). Assume an average molecular weight of 200.
8. One mole of an ideal monatomic gas is taken along a straight line path in the P-V diagram from the initial state $P_i = 1$ atm, $V_i = 2 \times 10^4\ \text{cm}^3$ to the final state $P_f = 2$ atm, $V_f = 3 \times 10^4\ \text{cm}^3$.
   (a) What are the initial and final values for the temperature?
   (b) How much work is carried out on the gas in this process?
   (c) How much heat must be added to the gas in this process?

*9. Find a path, consisting partially of an isobaric process and partially of a constant-volume process, that connects the two states of Problem 8 and involves no net heat loss.

10. Consider, in Figure 16-16, μ moles of a monatomic ideal gas taken from point 1 on the T_1-isotherm to point 3 on the T_2-isotherm along the path 1-2-3. In terms of the parameters shown, calculate (a) the energy change of the gas and (b) the heat that must be added to it in this process.

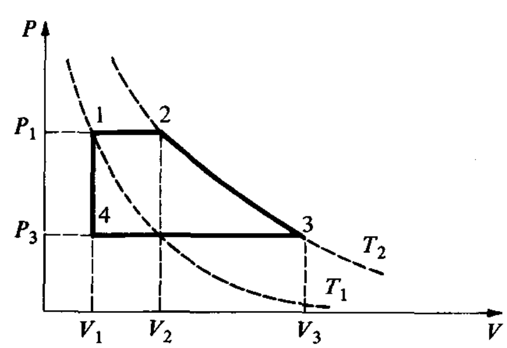


Figure 16-16

11. For the situation described in Problem 10, calculate the heat Q that must be added to the gas if it undergoes the process characterized by the rectilinear path 1-4-3.

12. One mole of an ideal monatomic gas is in the state corresponding to point A in Figure 16-17 and is taken along the path ABC in a certain process.
    (a) Calculate the temperature of the gas when it is at points A, B, and C.
    (b) How much work is carried out on the gas when it goes from A to B?
    (c) How much heat is added to the gas along the path BC?
    (d) What is the energy change of the gas during the process corresponding to the path ABC?

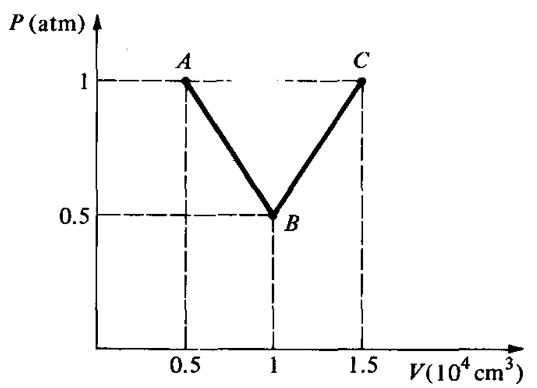


Figure 16-17

13. A meteorological balloon has a volume of 2 cubic meters at sea level and contains 8 moles of He at a pressure of 1.1 atm. After it has risen to an elevation of 5 km, its volume is 50 percent greater and the gas pressure has dropped to 0.6 atm.
    (a) What are the initial and final temperatures of the gas?
    (b) Assuming that, during the ascent, the change in volume of the gas is always proportional to the change in pressure, calculate the work carried out on the gas.
    (c) How much heat is absorbed by the gas in this process?
14. A gas is taken around the cycle $ABCDA$ in Figure 16-18.
    (a) How much work is carried out on the gas?

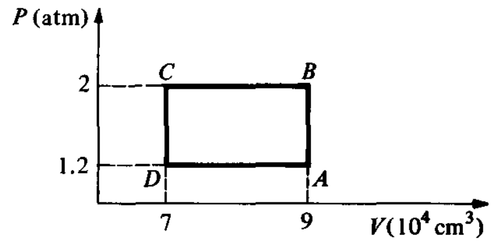


Figure 16-18

    (b) How much heat must be added to the gas in this process?
    (c) Repeat (a) and (b) if the cycle is traversed in the order $BADCB$?
15. Consider half a mole of an ideal monatomic gas confined to a cylinder of cross-sectional area 100 cm² and suppose that at one end there is a piston which is kept in place (as shown in Figure 16-19) by a spring of constant $k = 10^3$ N/m. Assume that originally the gas has a volume 10^4 cm³ and is at a temperature of 0°C.
    (a) What is the gas pressure?
    (b) By how much has the spring been compressed from its equilibrium length?

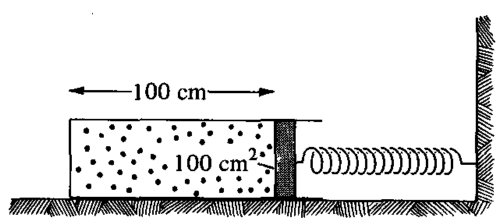


Figure 16-19

16. Consider again the physical situation in Figure 16-19 and suppose that the spring has its equilibrium length. Assume that $k = 10^3$ N/m, that the confined gas is ideal, monatomic, and at a temperature of 0°C.
    (a) How many moles of gas are in the container?
    (b) If the number of moles of gas is increased by 10 percent, by what amount is the spring compressed, assuming that the temperature does not change?
    (c) What is the gas pressure in (b)?

*17. An insulated cylinder closed at both ends contains an insulating piston, which is free to move on frictionless bearings and divides the cylinder into two compartments. Originally each compartment, call them A and B, has a volume $5.0 \times 10^4\,\text{cm}^3$ and contains a monatomic ideal gas at a temperature of 0°C and at a pressure of 1 atm. Suppose now that a certain amount of heat Q is slowly added to A so that the final pressure of the gas in B is 3 atm.
   (a) How many moles of gas are there in each compartment?
   (b) What is the final volume of the gas in B? (*Hint*: Why is the gas in B undergoing an adiabatic process?)
   (c) What are the final temperatures of the two gases?
   (d) How much heat was added to A?

18. An insulated cylinder is closed at one end and contains a movable insulated piston at the opposite end and a movable one in the middle. Suppose, as shown in Figure 16-20, that each compartment contains 1 mole of an ideal monatomic gas at a pressure P_0 and volume V_0. If the outer piston is slowly moved to the left so that the total volume of the gases becomes $(3/2)V_0$:
   (a) What are the final volumes of the gases in the two compartments?
   (b) What are the final pressures of the two gases?
   (c) How much work has been carried out by the external agent?

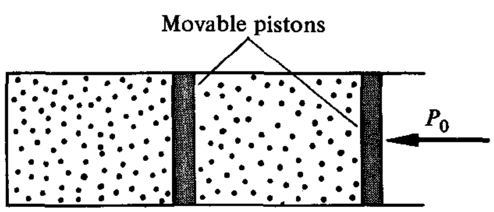


Figure 16-20

19. Suppose that the piston separating the two gases in Problem 18 is diathermal and fixed, and that the rise in temperature as a result of the compression is 10 K.
   (a) What is the final pressure of the gas on the left?
   (b) What is the final pressure of the gas on the right?
   (c) How much heat is transmitted to the gas on the left?
   (d) How much work has been carried out on the system?

20. One mole of an ideal monatomic gas is compressed adiabatically from an initial volume V_0 and pressure P_0 to a final volume V_1. In terms of the parameters P_0, V_0, and V_1, calculate:
   (a) The work carried out on the gas.
   (b) The energy change of the gas.
   (c) The heat added to the gas as predicted by use of your results to (a) and (b) and the first law.

21. Making use of the definition for the calorie and the Btu, the fact that 454 grams corresponds to 1 lb, and that there are 1.8°F in each degree Celsius, derive (16-5).

22. The energy E of a diatomic ideal gas of μ moles is

$$E = \frac{5}{2}\mu RT$$

where T is the absolute temperature of the gas. (a) Explain, in physical terms, why this energy is larger than that for a monatomic gas at the same temperature. (b) Show that for this diatomic gas, C_V and C_P are given by

$$C_V = \frac{5}{2}R \qquad C_P = \frac{7}{2}R$$

23. Show that in an adiabatic process the volume and pressure of a diatomic gas are related by

$$PV^{7/5} = \text{constant}$$

***24.** According to Debye, C_V for a crystalline solid at temperature T is

$$C_V = 9R\left[\frac{T}{\theta_D}\right]^3 \int_0^{\theta_D/T} dx \frac{x^4 e^x}{(e^x - 1)^2}$$

with θ_D the Debye temperature of the given solid. Some typical values are shown in the tabulation below.

(**a**) Show that for high temperatures, $T \gg \theta_D$, $C_V = 3R$.

(**b**) Show that for very low temperatures, $T \ll \theta_D$, $C_V \propto T^3$.

Substance	Al	Ag	Cr	Fe	Pb
θ_D (K)	428	225	630	467	105

25. Suppose 10 grams of steam at 100°C are forced to condense in 1 kg of water originally at 0°C. Calculate the final temperature of the water, assuming that the heat of vaporization of water is 540 cal/g.

26. A 5-gram cube of ice at 0°C is dropped into 20 grams of water at an original temperature of 80°C. Using the data in Table 16-5, calculate the final temperature of the water.

27. A steam generator is capable of taking 1 kg of water at 25°C and converting it, at a pressure of 1 atm, to steam at 100°C. Neglecting losses, how much energy does the water require to achieve this conversion?

28. Calculate the heat required to raise a 2-kg block of (**a**) silver; (**b**) iron; and (**c**) lead by 10°C. Assume in each case that $C_V = 3R$.

29. Assuming that for lead and copper $C_V = 3R$, calculate the final temperature of the system if a 1-kg chunk of lead at 800°C is dropped into 500 grams of water contained in a 300-gram copper bowl originally at 0°C.

30. Calculate the mininum mass of copper at 700°C that must be placed into 300 grams of water at 20°C so that all of the water becomes vaporized to steam at a pressure of 1 atm.

31. A 1-mole piece of lead at 0°C is placed in a container of liquid nitrogen that is at its boiling point of − 196°C. Using the data in Table 16-5 and assuming $C_V = 3R$ for lead, calculate the amount of nitrogen that is boiled off.

†32. Making use of the ideas of statistical mechanics and a very simple form for the intermolecular interaction, Van der Waals derived the equation of state

$$\left[P + a\left(\frac{\mu}{V}\right)^2\right][V - \mu b] = \mu RT$$

where R, P, V, T, and μ have their usual meanings, and where a and b are two positive constants.

(**a**) Show that for low densities, $\mu/V \to 0$, this reduces to the equation of state for the ideal gas.

(**b**) Show by use of (16-17) that, associated with this equation of state, there is a critical point characterized by the critical-point parameters

$$V_c = 3\mu b \qquad T_c = \frac{8a}{27Rb}$$

and find the value for P_c.

†33. Show that if the pressure, volume, and temperature in the Van der Waals equation of state in Problem 32 are expressed in units of their critical values, namely P_c, V_c, and T_c, then this equation of state may be written

$$p = \frac{8t}{3v - 1} - \frac{3}{v^2}$$

where $p = P/P_c$, $v = V/V_c$, and $t = T/T_c$. The fact that this equation correctly describes the thermodynamic behavior of many gases near their critical points is known as the *law of corresponding states.*

†34. Make a plot of the equation of states; that is, p versus v in Problem 33: (a) for $t > 1$; (b) for $t = 1$; and (c) for $t < 1$. Discuss the meaning of the unphysical parts of this isotherm.

35. A block of aluminum has a thickness of 1.0 cm and a cross-sectional area of 10^5 cm^2. If one side is maintained at a temperature of 30°C and the other at 80°C, calculate the rate at which heat is transported across the block. Use the value for k in Table 16-3.

36. The surface area of a house is 650 m^2. Assuming that the walls have an average thickness of 10 cm and consist of a material for which $k = 10^{-4}$ cal/cm-s-K, calculate the rate at which heat must be produced inside the house to maintain a temperature difference of 20 K with respect to the outside.

37. Suppose that the air just above the ice on a frozen river is at 10°F, that the ice layer is 10 cm thick, and that the water just below the ice is at the temperature of 32°F.
   (a) Calculate the upward rate of heat flow per unit area through the ice. Use the value $k = 5.2 \times 10^{-3}$ cal/s-cm-K for the thermal conductivity of ice.
   (b) Using the fact that the heat of fusion of ice is 80 cal/g, calculate the rate at which the ice becomes thicker.

38. Consider two blocks of material each of cross-sectional area A, of lengths l_1 and l_2, respectively, and of thermal conductivities k_1 and k_2. If, as shown in Figure 16-21, the left face is maintained at the temperature T_2 and the right-hand face is kept at the temperature T_1:
   (a) Show that the rate of heat flow $\Delta Q/\Delta t$ is

$$\frac{\Delta Q}{\Delta t} = A(T_2 - T_1)\left(\frac{l_1}{k_1} + \frac{l_2}{k_2}\right)^{-1}$$

   (b) Calculate the temperature at the interface between the two slabs and at all other points along the direction of heat flow.

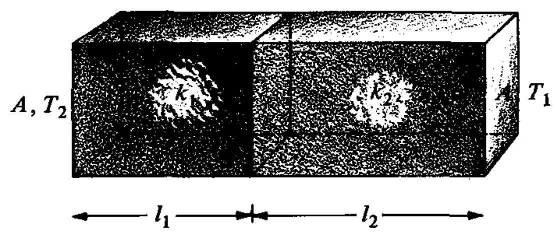


Figure 16-21

39. By analogy to the derivation for the heat flow in Example 16-11:
   (a) Show that the heat flow between the inner and outer surfaces of a hollow cylinder of inner and outer radii R_1 and R_2, respectively, and of length L is

$$\frac{\Delta Q}{\Delta t} = (T_2 - T_1)\frac{2\pi L k}{\ln (R_1/R_2)}$$

   (b) Calculate the temperature at a distance r from the axis of the cylinder.

17 Entropy and the second law of thermodynamics

The phenomena of the production of motion by heat has not been considered from a sufficiently general point of view. ... It is necessary to establish principles applicable not only to steam engines but to all imaginable heat engines, whatever the working substance, and whatever the method by which it is operated.

SADI CARNOT

Science may collect statistics and make charts. But its predictions are... but past history reversed.

JOHN DEWEY

17-1 Introduction

There are many physical processes, which, although consistent with the laws of mechanics, and thus of the first law, are nevertheless never observed to take place. The flow of heat from a cold body to a warmer one with which it is in contact is one example of such a process. Another is the spontaneous collapse of a gas to a volume smaller than that of its confining vessel. We call processes of this type, in which a physical system tends toward a nonequilibrium state, *unnatural*. The physical processes normally observed are called *natural*, and for these the system of interest invariably evolves toward a state

of thermal equilibrium. Typical of a natural process is the flow of heat from a hot body to a cooler one with which it is in contact.

By way of illustration, in Table 17-1 we list in column A a number of natural processes. Alongside each one in column B is listed the associated unnatural process, which would be observed if a motion picture film of the natural process were viewed when played backward. According to Newtonian mechanics, both the natural and the associated unnatural events are allowable. In other words, if we were to observe a motion picture film of any one of these processes, it is not possible to tell by use of Newtonian mechanics alone whether the film is being run forward or backward. Nevertheless, only those processes listed in column A are ever observed.

Table 17-1 Natural and unnatural processes

A *Natural processes*	B *Unnatural processes*
1. A gas is introduced into a vessel and expands to fill its volume.	1. A gas confined to a vessel collapses spontaneously to a volume smaller than that of the vessel.
2. An ice cube at 0°C is placed into a room at 20°C and melts to become water at 20°C while the room cools off slightly.	2. A small amount of water is in a room at 20°C and freezes to form an ice cube at 0°C while the room warms up slightly.
3. A block of wood sliding along a horizontal surface slows down and comes to rest while the surface becomes slightly warmer.	3. A block of wood is placed at rest on a hot horizontal surface, and starts to travel along the surface with increasing velocity while the surface becomes cooler.
4. A larger boulder rolls toward a house, strikes it, and reduces it to rubble.	4. A large boulder rolls through a pile of rubble and a house springs up.

Let us consider the third process listed in the table in more detail. Figure 17-1 shows the natural process of a block sliding initially at some velocity $\mathbf{v}_0$ over a rough, horizontal surface. Eventually it comes to rest and the surfaces heat up. This is a typical physical situation and we can readily understand the mechanism underlying this motion in microscopic terms. As the block slides

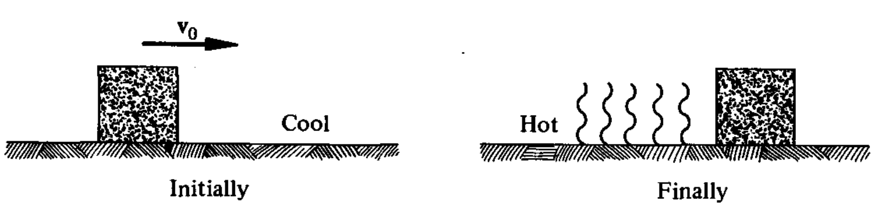


Figure 17-1

along the surface, its molecules and those comprising the surface collide with each other, and as a result their kinetic energies tend to increase. As we saw in connection with our studies of temperature, this increase in the thermal motion of the molecules manifests itself macroscopically as a rise in temperature. Thus, consistent with observation, the kinetic energy of the block steadily decreases while the thermal motions of the molecules involved increase.

Let us now contrast this familiar behavior with the related but *unnatural* process in Figure 17-2. This time, a block which is initially at rest on a heated horizontal surface starts to move spontaneously and acquires some final velocity $\mathbf{v}_0$ while the surfaces cool off. No one has ever observed this event!

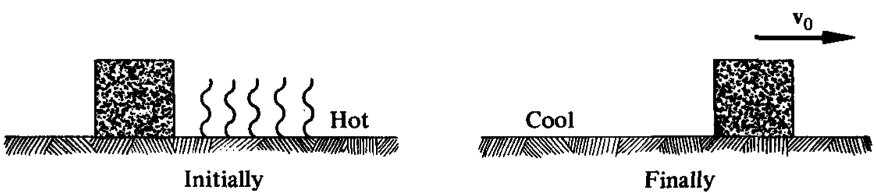


Figure 17-2

However, and this is the important point, the possibility of its taking place is *not* inconsistent with the laws of mechanics. Specifically, if the thermal motions of the molecules of the surface were just right so that they struck the block in the appropriate way to cause it to travel as shown in the figure, then this situation might very well take place. And, from the fact that it does not, we conclude that it must be highly improbable for the molecules comprising the two surfaces ever to acquire the necessary highly ordered motion. More generally, it is this inability of the molecules of ordinary matter to undergo highly organized collective behavior that makes unnatural processes such a rarity, or more properly, an impossibility.

A question that arises in this connection is whether there is some general physical principle or law which can be used to distinguish natural from unnatural processes. The answer to this question is affirmative. The physical law by means of which it is possible to predict *a priori* whether any given process is natural or unnatural is known as the *second law of thermodynamics.* The purpose of this chapter is to study this law.

17-2 Entropy

Consider a physical system of some type that is in thermodynamic equilibrium at temperature T. Suppose, as in Figure 17-3a, that it is subjected to a certain process, which involves adding to it an infinitesimal amount of heat ΔQ and carrying out on it a very small amount of work ΔW. According to the first law, after thermal equilibrium is reestablished, the energy of the system

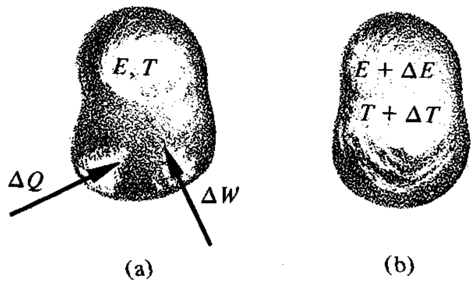


Figure 17-3

has changed by an amount ΔE, where

$$\Delta E = \Delta Q + \Delta W \tag{17-1}$$

In terms of these quantities, we now define the entropy S of this system by associating with the given infinitesimal process an entropy change ΔS where

$$\Delta S = \frac{\Delta Q}{T} \qquad \text{(infinitesimal process)} \tag{17-2}$$

Note that the process has been assumed to be infinitesimal and, in particular, that ΔQ is sufficiently small so that the absolute temperature T of the system does not change significantly during the process.

There are a number of properties of the entropy S that follow from the definition in (17-2) and deserve particular emphasis.

1. The entropy S of a system is defined only up to the addition of an arbitrary constant. Just as for energy, only differences in entropy will have meaning for us.
2. The change in entropy ΔS of a system undergoing an infinitesimal process can be positive, or negative (or zero) depending on the sign of ΔQ.
3. Since for an adiabatic infinitesimal process $\Delta Q = 0$, it follows that the change in entropy of a system undergoing such a process vanishes. Hence, we often refer to such a process as *isentropic*, which means constant entropy.
4. Just as is the energy E of a thermodynamic system, the entropy S is defined only for equilibrium states. We have *not* defined entropy for systems not in thermal equilibrium.

For a finite process for which the change in temperature of the system cannot be neglected, the entropy may be calculated in the following way. Think of the process as being carried out sequentially in a series of infinitesimal steps, so that the net change of entropy is the sum of the entropy changes for each infinitesimal process. It follows that the change in entropy ΔS for the complete process is

$$\Delta S = \sum \frac{\Delta Q_i}{T_i} \tag{17-3}$$

where ΔQ_i is the infinitesimal heat added to the system at the ith step when its temperature is T_i. If, for example, heat is added to μ moles of a system so that its temperature changes from an initial value T_1 to a final value T_2, then, assuming the process takes place at constant volume, the entropy change is

$$\Delta S = \int_{T_1}^{T_2} \mu C_V \frac{dT}{T} \tag{17-4}$$

where C_V is the molar heat capacity of the substance at constant volume and where the sum in (17-3) has been replaced by an appropriate integral.

As a special case, let us calculate the change in entropy of 1 gram of ice at 0°C when it melts to water at the same temperature. Since the heat of fusion of ice is 80 cal/g, it follows that 80 cal of heat is required to melt the ice. Therefore, since the temperature does not change in this process the result is

$$\Delta S = \frac{\Delta Q}{T} = \frac{80 \text{ cal}}{273 \text{ K}} = 0.29 \text{ cal/K}$$

Finally, there is the extremely important and crucial property of entropy that follows directly from (17-2) and experiment. This property is that entropy is a *state function.* Consider a physical system in some initial equilibrium state i, and imagine it being subjected to a variety of physical processes in each of which it starts out in the same state i and winds up in a specific final state f. If Q is the heat added to the system in any one of these processes and W is the corresponding work carried out, then, by the first law,

$$E_f - E_i = Q + W$$

where, even though for each process both Q and W will in general be different, their sum must be the constant difference $(E_f - E_i)$. If we now calculate the entropy change ΔS associated with any one of these processes we find that ΔS is also the same for each of them. In other words, regardless of what process is used to take the system from an initial equilibrium state i to a final equilibrium state f, the *entropy change is the same for all of them.* Thus, just as is the energy E, the entropy S of a system is also a state function, and the change in entropy for any process may be written

$$\Delta S = S_f - S_i \tag{17-5}$$

where S_f and S_i are the entropies of the final and initial states respectively.

17-3 Entropy calculations

In this section, we use the definition for entropy to calculate the entropy changes associated with a number of very simple processes.

Example 17-1 Calculate the entropy change associated with the boiling of 5 grams of water at 100°C to steam at the same temperature. What is the entropy change

associated with the reverse process—that is, of the condensation of 5 grams of steam at 100°C to water at the same temperature?

Solution Since the temperature of the water is constant while it boils, it follows, by use of (17-2) and the fact that the heat of vaporization of water at 100°C is 540 cal/g, that the entropy change ΔS is

$$\Delta S = \frac{\Delta Q}{T} = \frac{5\text{ g} \times 540\text{ cal/g}}{373\text{ K}} = 7.2\text{ cal/K}$$

The fact that ΔS is positive is consistent with the fact that heat must be added to water to cause it to boil, so that $\Delta Q > 0$.

By contrast, when steam condenses it is necessary to extract 540 cal/g of heat. Thus the change in entropy of 5 grams of steam in condensing to water at 100°C is $\Delta S = -7.2$ cal/K since we need only repeat the above calculation, replacing ΔQ by its negative. •

Example 17-2 A certain amount of heat ΔQ is added to μ moles of an ideal monatomic gas in a process in which it is kept at constant volume. If the initial temperature of the gas is T_0, calculate:

(a) Its final temperature.
(b) The entropy change of the gas.

Solution

(a) Since the volume of the gas is kept constant, we may use the formula for the molar specific heat at constant volume $C_V = \frac{3}{2}R$. Hence

$$\Delta Q = \mu C_V \,\Delta T = \frac{3}{2}\mu R(T_f - T_0)$$

where T_f is the sought-for final temperature. Solving for T_f, we obtain

$$T_f = \frac{2\,\Delta Q}{3\mu R} + T_0$$

(b) According to (17-4), the entropy change ΔS is

$$\Delta S = \int_{T_0}^{T_f} \frac{\mu C_V\, dT}{T} = \frac{3}{2}\mu R \int_{T_0}^{T_f} \frac{dT}{T}$$
$$= \frac{3}{2}\mu R \ln \frac{T_f}{T_0} = \frac{3}{2}\mu R \ln\left(1 + \frac{2\Delta Q}{3\mu R T_0}\right)$$

where the final equality follows by use of the result of (a). Since $\ln x$ is positive for $x > 1$ and negative for $x < 1$, the entropy change ΔS is positive when ΔQ is positive and negative when heat is extracted from the gas.

Example 17-3 One mole of water at 280 K is poured into a vessel of negligible heat capacity, which contains 1 mole of water originally at 320 K.

(a) What is the final temperature of the mixture?
(b) Calculate the entropy change ΔS_c of the originally cold water.
(c) Calculate the entropy change ΔS_h of the other mole of water.

Solution

(a) By symmetry, or by use of the known specific heats, the final temperature T_f of the mixture must be the arithmetic mean of the two:

$$T_f = \frac{1}{2}(280\text{ K} + 320\text{ K}) = 300\text{ K}$$

(b) Using the fact that for water $C_V = 18$ cal/mole-K, we find by use of (17-4) that

$$\Delta S_c = \int_{280\text{K}}^{300\text{K}} \mu C_V \frac{dT}{T} = \left(\frac{18\text{ cal}}{\text{mole-K}}\right) \times 1\text{ mole} \times \ln\frac{300}{280}$$
$$= 18 \times 0.069\text{ cal/K} = 1.24\text{ cal/K}$$

(c) In the same way we obtain

$$\Delta S_h = \int_{320\text{K}}^{300\text{K}} \mu C_V \frac{dT}{T} = \left(\frac{18\text{ cal}}{\text{mole-K}}\right) \times 1\text{ mole} \times \ln\frac{300}{320}$$
$$= -18 \times 0.0645\text{ cal/K} = -1.16\text{ cal/K}$$

Note that the sum $(\Delta S_c + \Delta S_h)$ is positive. As we shall see later, in all natural processes the change in entropy of all systems involved can never be negative.

Example 17-4 Consider again the physical situation described in Example 15-9 of an ideal gas in contact with a heat reservoir at temperature T_0 and suddenly compressed by a very large force F from an initial volume Al to a final volume $Al/2$. After thermal equilibrium is restored:

(a) What is the entropy change ΔS_R of the heat reservoir?
(b) What is the entropy change ΔS_g of the gas?

Solution

(a) Since the final temperature of the gas is the same as its initial value T_0, and since the energy of an ideal gas depends only on temperature, it follows that the change in energy ΔE of the gas is zero in this process. Moreover, as we saw in Example 15-9, the work carried out on the gas in this process is $Fl/2$. Hence, by the first law, the heat ΔQ added to the gas is

$$\Delta Q = -W = -\frac{Fl}{2}$$

The minus sign reflects the fact that heat is *extracted* from the gas.

Now this heat is absorbed by the heat reservoir during the time that the gas relaxes to thermal equilibrium. Since the temperature of the heat reservoir does not change when it absorbs ordinary amounts of heat, it follows from (17-3) that the entropy change ΔS_R of the heat reservoir is

$$\Delta S_R = \frac{-\Delta Q}{T} = \frac{Fl}{2T_0}$$

since $+Fl/2$ is the heat gained by the reservoir.

(b) To calculate the entropy change of the gas, we make use of the fact that the entropy change of a system undergoing some process depends only on the difference between the entropies of the initial and final states and is independent of the process

involved. In particular, we may calculate the entropy change of the gas by subjecting it to an isothermal process at temperature T_0 from an initial volume Al to a final one $Al/2$. According to (15-24), the work W carried out in such a process is

$$\mu R T_0 \ln\left(\frac{Al}{Al/2}\right) = \mu R T_0 \ln 2$$

Since the change in energy of an ideal gas vanishes for any isothermal process, it follows that

$$\Delta S_g = \frac{\Delta Q}{T_0} = -\frac{W}{T_0} = -\frac{\mu R T_0 \ln 2}{T_0} = -\mu R \ln 2$$

Just as in the preceding example, since by hypothesis the quantity F/A is much greater than the gas pressure, the sum of the entropy changes of the two systems is positive in this case as well.

17-4 Reversible and irreversible processes

In addition to natural and unnatural processes, for many purposes it is convenient to think of a third class of processes which falls somewhere between these two. These are known as *reversible processes.* As the name implies, a reversible process is one which can go in either direction. By definition, both natural and unnatural processes are *irreversible.*

More formally, we define a reversible process as one in which the system of interest remains in thermal equilibrium throughout. Or, equivalently, we say that a system undergoes a reversible process if it goes from a fixed initial state to a final state through a continuous sequence of equilibrium states. For example, if the system is a gas or a liquid, then for a reversible process it may be uniquely characterized by its volume and temperature alone at each stage of this process. It is apparent that, strictly speaking, reversible processes do not occur in nature any more than do unnatural ones. However, it is possible to represent to arbitrarily high precision any reversible process by a sequence of irreversible ones. Only in this highly theoretical way, then, can we say that reversible processes can take place.

To contrast the distinction between reversible and irreversible processes, consider the special case of the isothermal compression of an ideal gas. Figure 17-4 represents in a P-V diagram an ideal gas undergoing a reversible process involving its isothermal compression from an initial volume V_0 to a final one $V_0/2$. Since the process is reversible, the state of the system may be represented at *all* intermediate stages by a unique volume V and pressure P and by its temperature T_0. Thus, as shown in the figure, there is a continuous line connecting the initial point 1 to the final point 2. Let us now compare this idealized reversible process with the corresponding one as it might be carried out in the laboratory.

Consider an ideal gas in contact with a heat reservoir at temperature T_0 and suppose that first we compress it very quickly from an initial volume V_0 to a

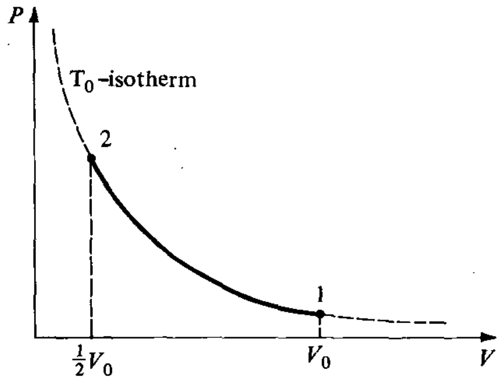


Figure 17-4

final one $V_0/2$. Figure 17-5a shows the appropriate P-V diagram. The gas starts out in thermal equilibrium at the point 1 on the T_0-isotherm corresponding to a volume V_0, but during the compression and immediately afterward it is *not* in thermal equilibrium. Thus its state *cannot* be represented on the P-V diagram during these intermediate stages. Ultimately, after enough heat has been extracted from the gas and the density inhomogeneities have damped out, it will return to a state of thermal equilibrium at the reservoir temperature T_0 and at a final volume $V_0/2$. At this point, it can again be represented on the P-V diagram as point 2 in Figure 17-5a. Note the important point that while the gas is in an intermediate state, it is not in thermal equilibrium and thus cannot be characterized by its volume and temperature alone.

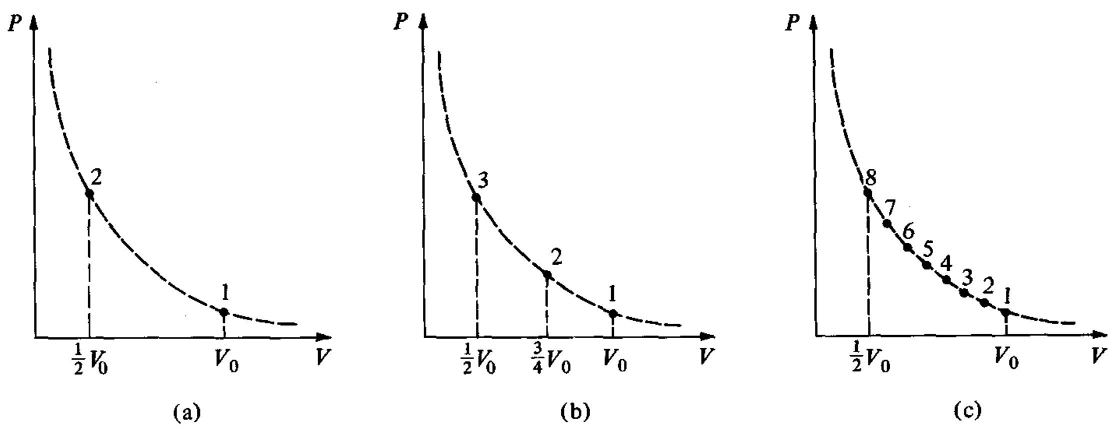


Figure 17-5

Let us repeat this process a second time, but now by compressing it first to an intermediate volume $3V_0/4$, allowing it to come to thermal equilibrium, and subsequently compressing it a second time to its final volume $V_0/2$. Using the above arguments, we may represent the situation this time as in Figure 17-5b by the three points labeled 1, 2, and 3 on the P-V isotherm. Similarly, Figure 17-5c shows the more complicated situation in which the gas is compressed in seven stages—being allowed to come to equilibrium after each intermediate compression. There are now eight points on the isotherm since the gas will be

in equilibrium at eight different stages. Proceeding in this way then, as we subdivide the act of compressing the gas into smaller and smaller steps, we approximate more and more closely the points on the continuous isotherm for the reversible process in Figure 17-4. And it is in this way that the process of compressing a gas can be thought to be reversible to arbitrarily high precision.

It is important to keep in mind that, strictly speaking, no process is truly reversible. Nevertheless, the concept of a reversible process is very fruitful and is in practice a very useful limiting case.

17-5 The second law of thermodynamics

Consider two arbitrary, *isolated* thermodynamic systems A and B which are initially in thermal equilibrium and in states characterized by the respective entropies S_A and S_B. Let us connect the two systems in such a way that they can exchange heat and work. For example, the systems might be two gases confined to an insulating container and separated by a diathermal, freely movable piston as in Figure 17-6. Or system A might be an ice cube, which is dropped into system B, a glass of water at room temperature. In general, both systems will change their states in some way until ultimately thermal equilibrium is reestablished at certain final values for their entropies; let us call them $(S_A + \Delta S_A)$ and $(S_B + \Delta S_B)$, respectively.

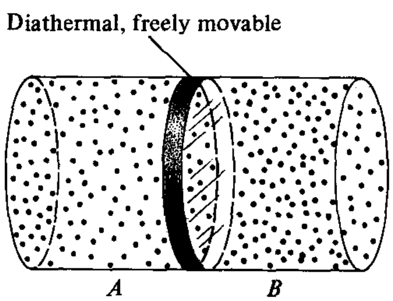


Figure 17-6

Now, depending on the specifics of the systems and the processes involved, these changes in entropy ΔS_A and ΔS_B can assume both positive and negative values. However, and this is crucial, experiment shows that regardless of the detailed nature of A and B and regardless of their initial states, the sum $(\Delta S_A + \Delta S_B)$ of the changes of their entropies is very severely restricted. This restriction is

$$\Delta S_A + \Delta S_B \geqslant 0 \tag{17-6}$$

with the equality sign prevailing only if both processes are reversible. In other words, for all processes that have ever been observed, the quantity $(\Delta S_A + \Delta S_B)$ has been found to be positive. For reversible processes, on the other hand, it vanishes, and if it were possible to observe unnatural processes we would obtain a negative value for this sum. Although for any given

processes either ΔS_A or ΔS_B may by itself be negative, the entropy change of the other system will always be sufficiently positive so that their sum will be positive. The fact that for all observed processes the inequality in (17-6) holds is known as the *second law of thermodynamics.*

Besides this formulation of the second law, there are a number of alternate ways of stating this law which are also of some interest. Two of these which are particularly noteworthy are the statements of the second law by Rudolph J. E. Clausius (1822–1888) and by Lord Kelvin (1824–1907). They are:

Clausius: *It is impossible to subject a system to a process whose only effect is to convey heat from a cooler to a hotter body.*

Kelvin: *It is impossible to have a process whose only effect is the absorption of heat from a heat reservoir at a single temperature throughout and the conversion of this heat completely into mechanical work.*

To make plausible the Clausius statement, suppose it were possible to extract heat Q (>0) from a reservoir at temperature T_1 and add it to a hotter one at T_2. Since by hypothesis all other systems involved in this heat transfer return to their original states at the end of the process, it follows that the total entropy change ΔS is that of the reservoirs alone. According to (17-3), ΔS has the value

$$\Delta S = -\frac{Q}{T_1} + \frac{Q}{T_2} = Q\left[\frac{1}{T_2} - \frac{1}{T_1}\right]$$

But $T_1 < T_2$, and hence this is negative in violation of the second law.

Similarly, to make plausible the Kelvin statement, suppose that we could devise a way for extracting Q positive units of heat from a reservoir at temperature T and converting it completely to mechanical work without changing the state of any other system. Again, the total entropy change is that of the reservoir alone and has the negative value $-Q/T$. Hence, assuming the validity of the second law, the Kelvin statement follows. In a similar way, (17-6) can be derived from either the Clausius or the Kelvin statement of the second law. Hence all three of these forms of the law are equivalent.

Example 17-5 Suppose that the gases A and B in the two compartments of the vessel in Figure 17-6 are monatomic, have the same number of moles μ, and are initially at the temperatures T_A and T_B, respectively. Assuming that the partition between them is diathermal but does *not* move, calculate the change in entropy of the gases when thermal equilibrium is finally established.

Solution Since the number of molecules of each of the gases is the same, it follows by use of (15-18) that the temperature T when thermal equilibrium is established is

$$T = \frac{1}{2}(T_A + T_B)$$

To calculate the entropy change of each system let us make use of (17-4) and the fact that at constant volume the molar heat capacity C_V of an ideal gas is $3R/2$. The entropy change ΔS_A of the gas initially at temperature T_A is

$$\Delta S_A = \int_{T_A}^{T} \mu C_V \frac{dT}{T} = \frac{3}{2}\mu R \ln \frac{T}{T_A}$$

and for ΔS_B we find the same formula except with T_A replaced by T_B. Hence the total change in entropy is

$$\Delta S = \Delta S_A + \Delta S_B = \frac{3}{2}\mu R\left[\ln \frac{T}{T_A} + \ln \frac{T}{T_B}\right]$$

$$= \frac{3}{2}\mu R \ln \frac{T^2}{T_A T_B} = 3\mu R \ln\left(\frac{T_A + T_B}{2\sqrt{T_A T_B}}\right) > 0$$

where the third equality follows since $\ln x + \ln y = \ln xy$. The final inequality is then a consequence of the fact that the arithmetic mean $\frac{1}{2}(T_A + T_B)$ of two positive numbers cannot be less than their geometric mean $(T_A T_B)^{1/2}$. Consistent with the second law, the entropy increases.

Example 17-6 Suppose that 10 grams of water at 10°C are poured into a flask containing 50 grams of water originally at 80°C. Using the fact that the molar heat capacity of water is 18 cal/mole-K and neglecting the heat capacity of the container, calculate:

(a) The final temperature of the water.
(b) The entropy change of the water originally at 10°C.
(c) The entropy change of the water originally at 80°C.
(d) The total entropy change of the system.

Solution

(a) To find the final temperature T of the mixture we proceed as in Chapter 16 and write:

heat gained by the 10 grams of water = heat lost by the 50 grams of water

Using the fact that the molecular weight of water is 18, the heat gained by the 10 grams of water is

$$\mu C_V \, \Delta T = \left(\frac{10\text{ g}}{18\text{ g/mole}}\right) \times \left(\frac{18\text{ cal}}{\text{mole-K}}\right) \times (T - 283\text{ K})$$
$$= (10T - 2830\text{ K})\text{ cal/K}$$

with all temperatures expressed on the Kelvin scale. Similarly, the heat lost by the 50 grams of water is

$$(17{,}650\text{ K} - 50T)\text{ cal/K}$$

Equating these two we find that

$$17{,}650 - 50T = 10T - 2830$$

and this leads to

$$T = 341\text{ K} = 68°\text{C}$$

(b) In general, the entropy change ΔS of a system undergoing a constant-volume process from an initial temperature T_i to a final one T_f is

$$\Delta S = \int_{T_i}^{T_f} \mu C_V \frac{dT}{T} = \mu C_V \int_{T_i}^{T_f} \frac{dT}{T}$$
$$= \mu C_V \ln\left(\frac{T_f}{T_i}\right)$$

provided that its molar heat capacity C_V is constant.

Making use of this formula, for the entropy change ΔS_{10} of the 10 grams of water, we obtain

$$\Delta S_{10} = \mu C_V \ln\left(\frac{T_f}{T_i}\right) = \left(\frac{10\text{ g}}{18\text{ g/mole}}\right) \times \left(\frac{18\text{ cal}}{\text{mole-K}}\right) \times \ln\frac{341\text{ K}}{283\text{ K}}$$
$$= 1.86\text{ cal/K}$$

where in the final equality we have used the tabulated value $\ln 1.20 = 0.186$.

(c) In the same way, the entropy change ΔS_{50} of the 50 grams of water is

$$\Delta S_{50} = (50\text{ cal/K}) \times \ln\frac{341\text{ K}}{353\text{ K}}$$
$$= -1.73\text{ cal/K}$$

where we have used the tabulated value $\ln 0.966 = -0.0346$. The entropy change ΔS_{50} is negative here, thus reflecting the fact that heat is extracted from this water originally at 80°C.

(d) Combining these results, we obtain for the total entropy change

$$\Delta S = \Delta S_{10} + \Delta S_{50} = 1.86\text{ cal/K} - 1.73\text{ cal/K} = +0.13\text{ cal/K}$$

Fully in accordance with the second law, this is positive. Hence, this process is natural and irreversible.

17-6 The entropy of the ideal gas

The purpose of this section is to derive the formula

$$S = Nk\left[\frac{3}{2}\ln E + \ln V - \frac{5}{2}\ln N\right] \tag{17-7}$$

for the entropy S of a monatomic ideal gas of N molecules when its temperature is T and it occupies a volume V. Since the gas is ideal its energy E may be expressed in terms of T by

$$E = \frac{3}{2} NkT \tag{17-8}$$

To derive this formula for S, imagine subjecting a monatomic ideal gas to an infinitesimal reversible process in which its volume, energy, and temperature change by the respective infinitesimal amounts dV, dE, and dT. Since the

process is reversible, the work carried out on it is $-P\,dV$, where P is the gas pressure, and hence the heat added to it in this process is $dE-(-P\,dV)=(dE+P\,dV)$, according to the first law. Substituting into (17-2), we obtain, for the entropy change dS of the gas,

$$dS=\frac{dE}{T}+\frac{P}{T}dV \tag{17-9}$$

Let us now substitute for the factor $1/T$ in the first term on the right-hand side by use of (17-8) and for the factor P/T in the second term by use of the equation of state for the ideal gas in (15-10). There results

$$dS=\frac{3}{2}Nk\frac{dE}{E}+Nk\frac{dV}{V}$$

and this integrates to

$$S=\frac{3}{2}Nk\ \ln E+Nk\ \ln V+C \tag{17-10}$$

where C is a constant of integration.

Let us now turn to the problem of fixing this integration constant C in (17-10). Since the entropy of the system has been defined only up to an additive constant, it would appear to be simplest to set this constant to zero. However, as was first pointed out by J. Willard Gibbs (1839–1903), if we do this there arises a certain inconsistency in the formula for the entropy in (17-10). This is now known as the *Gibbs paradox.*

Consider, in Figure 17-7, the N molecules of a monatomic ideal gas at temperature T confined to a volume V. Imagine the vessel being divided into two equal parts by a fictitious surface (the dotted surface A in the figure) so that the parameters characterizing the gas in each half are $N/2$, $V/2$, and T. If E represents the total energy of the gas, then in accordance with (17-8) the energy of the gas in each half of the container is $E/2$. Hence, by use of (17-10), it follows that the entropies S_1 and S_2 in each half of the container are the same and have the value

$$S_1=S_2=\frac{3}{2}\frac{Nk}{2}\ln\frac{E}{2}+\frac{Nk}{2}\ln\frac{V}{2}+C$$

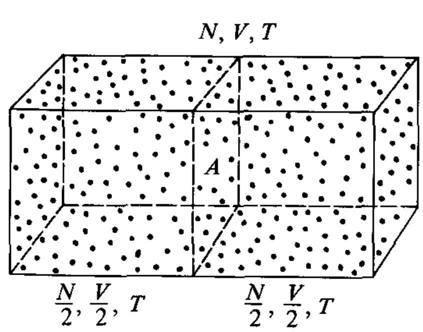


Figure 17-7

Thus on the one hand the total entropy $S = S_1 + S_2$ is

$$S = S_1 + S_2 = \frac{3Nk}{2} \ln \frac{E}{2} + Nk \ln \frac{V}{2} + 2C \tag{17-11}$$

while on the other the total entropy S must also be expressible in the form in (17-10). Since these two formulas for the same physical quantity differ for every choice for the constant C, we conclude that (17-10) cannot be correct without some modification.

Brief reflection of this seeming contradiction shows that—since the energy and the volume dependence of the entropy in (17-11) is the same as that in (17-10)—these formulas could be made consistent by allowing the constant C to depend on the number of particles, N. Indeed in the problems it is shown that with the choice

$$C = -\frac{5}{2} Nk \ln N \tag{17-12}$$

there is no longer a paradox. Thus we obtain the final form for the entropy S of an ideal gas in (17-7).

We emphasize that the entropy formula in (17-7) depends on the validity of the energy formula in (17-8) and thus applies only for a *monatomic* ideal gas. The corresponding formula for a diatomic gas, as shown in the problems, is

$$S = \frac{5}{2} Nk \ln E + Nk \ln V - \frac{7}{2} Nk \ln N \tag{17-13}$$

Example 17-7 One mole of a monatomic ideal gas confined to one compartment of an insulated container undergoes a free expansion from an initial volume V_0 to a final volume $2V_0$. Calculate the change in its entropy.

Solution In Chapter 13 we saw that in a free expansion the thermal velocity of an ideal gas remains the same. Hence it follows that in such an expansion the temperature, and thus the energy, of the gas are also not changed. Making use of (17-7) and the fact that $R = N_0 k$, we find for the initial and final values for the entropy of the gas S_i and S_f respectively,

$$S_i = \frac{3}{2} R \ln E + R \ln V_0 - \frac{5}{2} R \ln N_0$$

$$S_f = \frac{3}{2} R \ln E + R \ln 2V_0 - \frac{5}{2} R \ln N_0$$

where N_0 is Avogadro's number. The change in entropy is thus

$$\Delta S = S_f - S_i = R \ln 2$$

The fact that this change in entropy ΔS is positive shows that, consistent with experiment, this process is natural. Can you explain the source of this entropy increase in physical terms in light of (17-2) and the fact that no heat enters the gas from the outside?

Example 17-8 For a monatomic ideal gas, determine the functional dependence of the entropy on T and V.

Solution The substitution of (17-8) into (17-7), yields

$$S = Nk\left[\frac{3}{2}\ln T + \ln V - \ln N\right]$$

plus a constant term $Nk \ln(3k/2)$ which we have neglected.

Example 17-9 An insulated container consists of two compartments, each of volume V_0 and separated by a partition. Suppose that initially each compartment is occupied by the N atoms of a dilute gas each at temperature T_0. If the partition is removed, calculate the total change in entropy if:

(a) Both gases are helium.

(b) One gas is helium and the other is argon.

Solution

(a) Expressing the result of Example 17-8 in the form

$$S = Nk \ln\left[\frac{T_0^{3/2} V_0}{N}\right]$$

and noting that in the present case neither T_0 nor the ratio V_0/N changes as a result of removing the partition, we conclude that the entropy change is zero.

(b) For the case of two different gases we proceed as follows. The initial total entropy S_i is

$$S_i = Nk \ln\left[\frac{T_0^{3/2} V_0}{N}\right] + Nk \ln\left[\frac{T_0^{3/2} V_0}{N}\right]$$

where the first term represents the entropy of, say, the helium and the second, that of the argon. The final value S_f after the partition is removed is

$$S_f = Nk \ln\left[\frac{T_0^{3/2}\, 2V_0}{N}\right] + Nk \ln\left[\frac{T_0^{3/2}\, 2V_0}{N}\right]$$

since each gas occupies a final volume $2V_0$ and the temperature remains the same. Hence, by subtraction,

$$\Delta S = S_f - S_i = 2Nk \ln 2$$

This increase ΔS in entropy is known as the *entropy of mixing*. Physically it corresponds to the fact that the reverse of this process, the spontaneous separation of a mixture of two gases into its distinct components, is unnatural. This is in contrast to the above "mixing" of identical gases for which both the process and its reverse are natural. More generally, we find that we can associate decreasing order (or increasing disorder) with a rise in the entropy of any system. Thus the mixing of the two helium gases corresponds to no increase in disorder, and thus $\Delta S = 0$. On the other hand, if we mix two dissimilar gases there is a decrease in order, and hence the entropy increases.

17-7 Cyclic heat engines

In the remainder of this chapter we shall be concerned with certain thermodynamic systems known as *cyclic heat engines.* Our main goal will be to derive certain theorems originally proposed by Sadi Carnot (1796–1832), which played a vital role in the discovery and development of the second law. We shall conclude by making use of these ideas to define the *absolute* or *thermodynamic* temperature scale.

A *heat engine* is a thermodynamic system that is capable of absorbing energy in the form of heat from a source, such as a burning fuel, and converting part of this energy into mechanical work. The fact that not all of this absorbed heat can be converted to work is a consequence of the second law. A *cyclic heat engine* is a type of heat engine that operates in a cycle, and thus at the end of each cycle the thermodynamic system that defines the engine returns to its original state. Examples of cyclic heat engines abound in any industrial society. Typical are the combustion engine, such as that in an automobile; the steam engine; and the diesel engine. In the following we shall not go into the detailed mechanisms underlying the operation of cyclic heat engines; rather, we shall be concerned only with very general and overall aspects of such engines.

In general, a cyclic heat engine is a system that, during each cycle, extracts heat from a heat reservoir maintained at one temperature, converts part of this energy to useful work, and rejects the remainder to a heat reservoir at a lower temperature. For example, in the combustion engine the input heat comes at the higher temperature from the combustion of gasoline, and the rejected heat leaves the engine at the lower temperature of the exhaust. It is convenient to represent this operation of a cyclic engine diagrammatically as in Figure 17-8. Here, the cyclic engine operates between two heat reservoirs, one maintained at a temperature T_1 and the second at a lower temperature T_2. During each cycle, the engine absorbs a certain amount of heat, call it Q_1, at the higher-temperature reservoir, converts some of this heat to useful work W, and rejects heat Q_2 to the lower-temperature reservoir. In other words,

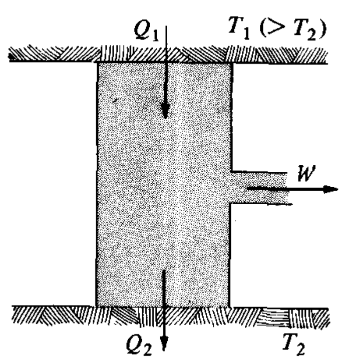


Figure 17-8

Figure 17-8 represents a heat engine, which absorbs during each cycle $(Q_1 - Q_2)$ units of heat and performs a certain amount of work W in the process.

If ΔE represents the change in the energy of the heat engine during each cycle, then, according to the first law,

$$\Delta E = Q_1 - Q_2 - W$$

since $-W$ represents the work carried out *on* the thermodynamic system constituting the engine. Now since the engine operates in a cycle and the system is in thermodynamic equilibrium at the beginning and end of each cycle, the energy change ΔE of the system vanishes. Accordingly, setting $\Delta E = 0$ we obtain the very fundamental relation

$$W = Q_1 - Q_2 \qquad \textbf{(17-14)}$$

which characterizes all cyclic engines.

A *refrigerator* is a cyclic heat engine that is run backward. In the operation of a refrigerator, during each cycle a certain amount of heat Q_2 is absorbed from a heat reservoir, work W is carried out by the external agent *on* the engine, and finally heat Q_1 is rejected to a second heat reservoir at a higher temperature. The operation of a refrigerator is represented schematically in Figure 17-9. In an ordinary household refrigerator, for example, the work W is normally carried out by an electric motor, which causes the transfer of heat from the low-temperature interior of the refrigerator to the generally hotter external environment. Just as for a cyclic heat engine, the energy change ΔE of a refrigerator must vanish during each cycle. Hence, by the first law,

$$0 = \Delta E = Q_2 - Q_1 + W$$

where we now have used the sign convention appropriate to the refrigerator in Figure 17-9. Solving for the work W we find that

$$W = Q_1 - Q_2 \qquad \textbf{(17-15)}$$

which is precisely the same as (17-14), but now the symbols must be interpreted as in Figure 17-9.

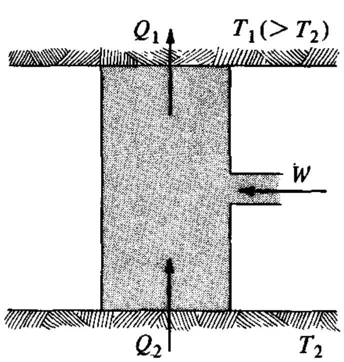


Figure 17-9

Of considerable theoretical importance is a certain idealized cyclic heat engine known as a *reversible* or a *Carnot* engine. A reversible engine is one whose material constituents are in a state of thermal equilibrium throughout the *entire* cycle. If, for example, the working substance of a reversible engine is a gas, then throughout the entire cycle the gas can be uniquely characterized by its volume and temperature alone. It should be noted that the cyclic heat engines we normally encounter are of the irreversible type, for which the thermodynamic system defining the engine is not in thermal equilibrium at some point during each cycle. A reversible engine undergoes no entropy change during its operation.

Finally, we define a *reversible refrigerator* as one for which the thermodynamic system constituting the refrigerator is in thermal equilibrium throughout its entire cycle. Equivalently, a reversible refrigerator is a reversible heat engine that is run backward.

17-8 The efficiency of heat engines

We define the efficiency η of a cyclic heat engine to be the ratio of the work W carried out by it during each cycle to the heat Q_1 it absorbs from the higher-temperature reservoir. If it were possible to convert all of this heat Q_1 to work, its efficiency would be unity. As we shall see below, however, an engine with this efficiency is not even theoretically possible.

In mathematical terms, the efficiency η of a cyclic engine is

$$\eta = \frac{W}{Q_1}$$

and substituting for W by use of (17-14) we obtain the equivalent formula

$$\eta = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1} \tag{17-16}$$

In Section 17-9 we shall demonstrate that, because of the second law, the heat Q_2 rejected per cycle at the lower temperature reservoir can never vanish. Hence, the efficiency η must satisfy the inequalities

$$0 \leq \eta < 1 \tag{17-17}$$

For typical cyclic engines, values for η vary in the range from 0.10 to 0.40. For example, for an efficient automobile engine η is approximately 0.20, while the efficiencies of diesel engines fall in the range from 0.35 to 0.40.

Example 17-10 A certain heat engine, which operates with an efficiency of 18 percent, absorbs 10^5 cal of heat per cycle from the high-temperature reservoir.

(a) How much heat is rejected per cycle at the low-temperature reservoir?

(b) How much work does the engine carry out during each cycle?

Solution In the notation of (17-14) and (17-16), the parameter values are $Q_1 = 10^5$ cal and $\eta = 0.18$.

(a) Solving (17-16) for Q_2 and inserting the given values, we obtain

$$Q_2 = (1-\eta)Q_1 = (1-0.18)\times 10^5 \text{ cal} = 8.2\times 10^4 \text{ cal}$$

(b) Substituting the values for Q_1 and Q_2 into (17-14), we find that

$$W = Q_1 - Q_2 = 10^5 \text{ cal} - 8.2\times 10^4 \text{ cal} = 1.8\times 10^4 \text{ cal}$$
$$= 7.5\times 10^4 \text{ J}$$

where the final equality follows by use of Table 16-1.

17-9 Carnot's theorems

In connection with his studies of cyclic heat engines, Carnot originated and developed the concept of a reversible engine. In his classic *Reflections on the Motive-Power of Heat,* which was published in 1824, he derived a number of very interesting properties of both reversible and irreversible heat engines. Of particular interest to us is his demonstration that the efficiencies of *all* reversible engines that operate between the same two heat reservoirs are the same. Hence this efficiency can depend only on the temperature of the reservoirs but not on any of the details of the reversible engines themselves. Also of interest is his proof that no ordinary engine operating between two reservoirs at fixed temperatures can have an efficiency exceeding that of a reversible engine operating between the same two reservoirs. Thus he established, once and for all, the maximum possible efficiency of all heat engines regardless of the ingenuity used in their design and construction.

Let us now state and prove the two key theorems that establish these properties.

Theorem I

> The efficiency of all reversible heat engines operating between two heat reservoirs at the fixed temperatures T_1 and T_2 is the same.

To prove this theorem, consider two reversible heat engines e and e', and suppose that when operating between two heat reservoirs at the temperatures T_1 and T_2 they have the respective efficiencies η and η'. Imagine now the compound engine consisting of e and e' operating together, as shown in Figure 17-10. Let the reversible engine e operate so that during each cycle it extracts heat Q_1 from the T_1-reservoir, rejects Q_2 units at the lower T_2-reservoir, and performs $W = Q_1 - Q_2$ units of work. At the same time let the reversible engine e' run backward as a refrigerator by making use of this work output W from e as input mechanical work. If e' therefore extracts Q_2'

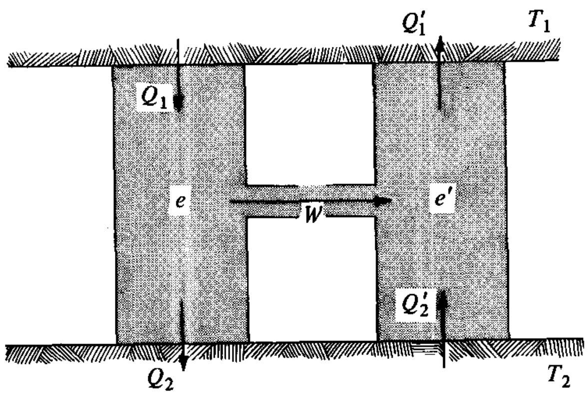


Figure 17-10

units of heat from the T_2-reservoir and rejects Q_1' at the other one, then, according to (17-14) and (17-15),

$$W = Q_1 - Q_2 = Q_1' - Q_2' \tag{17-18}$$

After the system has run through a complete cycle, the reversible engines e and e' will return to their initial states so that their entropy change is zero. On the other hand, the T_1-reservoir undergoes a change in entropy

$$\Delta S_1 = \frac{Q_1' - Q_1}{T_1}$$

since it absorbs $(Q_1' - Q_1)$ units of heat per cycle. Similarly, the entropy change of the lower temperature reservoir is

$$\Delta S_2 = \frac{Q_2 - Q_2'}{T_2} = -\frac{Q_1' - Q_1}{T_2}$$

where the second equality follows by use of (17-18). Hence the total entropy change of the system is

$$\Delta S = \Delta S_1 + \Delta S_2 = (Q_1' - Q_1)\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \tag{17-19}$$

Now, since $T_1 > T_2$, it follows from the second law that

$$Q_1' \leqslant Q_1 \tag{17-20}$$

and hence

$$\eta = \frac{Q_1 - Q_2}{Q_1} = \frac{Q_1' - Q_2'}{Q_1} \leqslant \frac{Q_1' - Q_2'}{Q_1'} = \eta' \tag{17-21}$$

In writing down the second equality we have made use of (17-18), and the inequality then follows from (17-20). The last equality is then a consequence of (17-16) and the fact that e' is a reversible engine.

Having thus established that e is no more efficient than is e', by reversing the role of the two engines we can by use of the same argument conclude that

the efficiency of e' cannot be less than that of e. Hence follows the sought-for relation

$$\eta = \eta' \qquad \textbf{(17-22)}$$

and this completes the proof that all reversible engines operating between the same two temperature reservoirs have the same efficiency.

Theorem II

> The efficiency η_i of any irreversible heat engine operating between two heat reservoirs at temperatures T_1 and T_2 is less than the efficiency η of any reversible engine operating between the same two temperatures.

To prove this theorem, consider again the situation in Figure 17-10. This time let e be an irreversible engine with efficiency η_i and let e' continue to represent a reversible engine driven backward as a reversible refrigerator by the output W of e. Proceeding as above, all of the steps up to (17-19) follow. Because of the fact that e is no longer reversible, the analogue (17-20) must be a strict inequality since, according to the second law for an irreversible process, the entropy must increase. Hence we conclude this time that

$$\eta_i < \eta' \qquad \textbf{(17-23)}$$

Moreover, since e is now irreversible, it is not possible to reverse the roles of the engines and conclude the opposite inequality. The validity of Theorem II is thus established.

†17-10 The Carnot cycle and the thermodynamic temperature scale

According to the first of Carnot's theorems the efficiencies of all reversible engines operating between two fixed temperatures T_1 and T_2 must be the same. In particular, this means that the efficiency of a reversible engine can be a function only of these temperatures T_1 and T_2 and cannot depend on the detailed nature of the engine. Hence, to determine the functional dependence of the efficiency η on T_1 and T_2, it suffices to evaluate η once and for all for any convenient reversible engine. The simplest thermodynamic system that we know of is the monatomic ideal gas; thus let us calculate η by making use of a reversible engine that uses, as a working substance, a dilute gas.

Consider the P-V diagram in Figure 17-11, which represents a monatomic ideal gas taken around a reversible cycle along the path *abcda*. In this diagram the paths *ab* and *cd* are portions of two isotherms at the respective temperatures T_1 and T_2, and the remaining two paths *bc* and *da* are adiabats. In this process, which is known as a *Carnot cycle*, the gas starts out at point *a*,

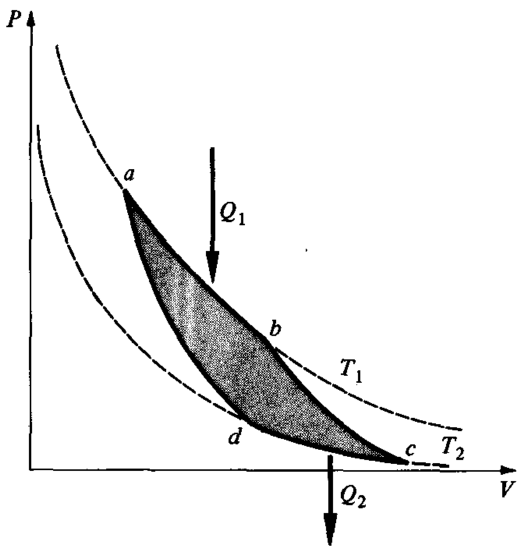


Figure 17-11

characterized by P_a, V_a, and T_1, and expands isothermally to point b, characterized by P_b, V_b, and T_1. In the second stage it expands adiabatically from b to point c, which is on the lower-temperature T_2-isotherm, and is characterized by P_c and V_c. During the third stage it is compressed isothermally at temperature T_2 until it reaches point d, characterized by P_d and V_d, and finally it is compressed adiabatically back to the original point a. Figure 17-12 shows these four parts of a Carnot cycle more graphically.

Now during the first stage, while the gas expands isothermally at temperature T_1, it absorbs a certain amount of heat Q_1 from the heat reservoir with which it is in contact. Similarly, during the third stage, while it is being compressed isothermally at temperature T_2, it rejects a certain amount of heat Q_2 to the lower-temperature reservoir at T_2. During the second and fourth stages, since the processes are adiabatic, no heat is extracted from or added to the gas. (See Example 16-4.) In Example 17-13, we shall establish that these heats Q_1 and Q_2 depend only on the temperatures of the reservoirs and have the ratio

$$\frac{Q_2}{Q_1} = \frac{T_2}{T_1} \tag{17-24}$$

Hence making use of this relation and the definition for efficiency in (17-16), we conclude:

The efficiency η of any reversible cyclic heat engine that operates between two heat reservoirs at temperatures T_1 and T_2 is

$$\eta = 1 - \frac{T_2}{T_1} \tag{17-25}$$

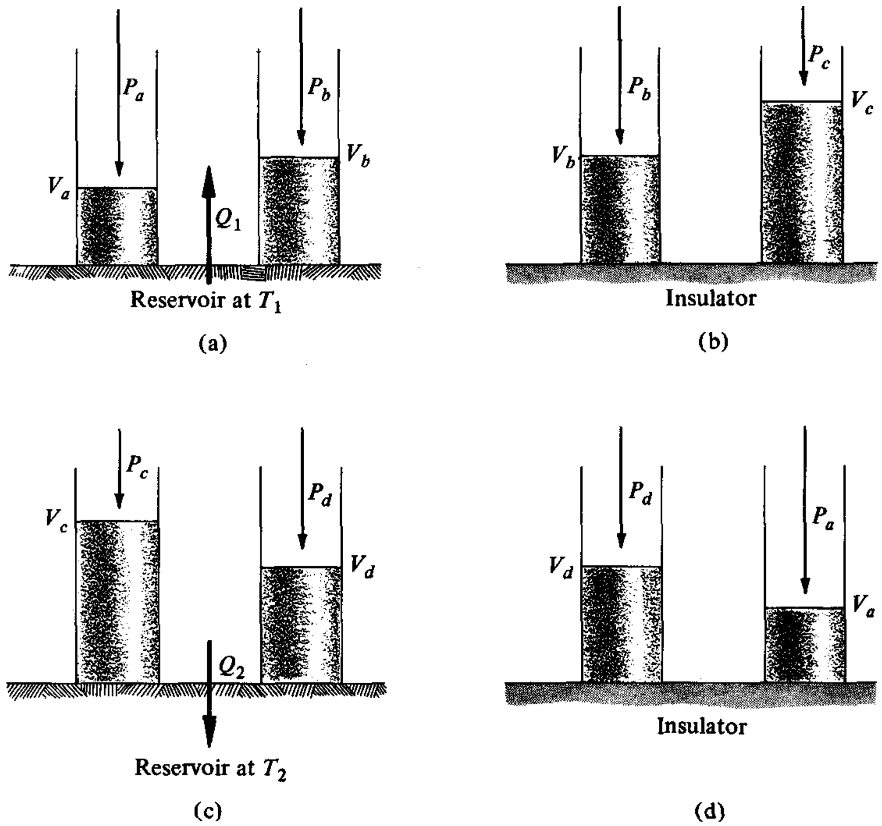


Figure 17-12

This very basic formula, when used in conjunction with Carnot's Theorem II, sets an upper limit to the efficiency of all conceivable cyclic heat engines operating between two given heat reservoirs.

Before considering in detail the required proof of (17-24), let us define the *absolute* or the *thermodynamic* temperature scale by its use.

Lord Kelvin was the first to suggest making use of these very remarkable results of Carnot, which led to (17-24) and (17-25), to define a thermodynamic temperature scale. To this end, let τ_1 and τ_2 be the respective temperatures, according to such a scale, of two heat reservoirs between which operates a reversible, cyclic heat engine. Bearing in mind (17-24), we impose the condition

$$\frac{\tau_2}{\tau_1} = \frac{Q_2}{Q_1} \tag{17-26}$$

with Q_1 representing the heat extracted at the higher-temperature reservoir and Q_2 that rejected at the lower. Since for any two heat reservoirs, the ratio Q_2/Q_1 can be measured directly in terms of the efficiency of a reversible engine operating between them, it follows that a unique value for the new temperature ratio τ_2/τ_1 can also be associated with these reservoirs. As for the constant-volume gas thermometer, let us fix the proportionality constant involved by assigning the value $\tau_t = 273.16$ to the temperature of a heat

reservoir when it is in thermal equilibrium with water at its triple point. With the constant of proportionality thus fixed, the temperature τ of any other heat reservoir may be obtained in accordance with the formula

$$\tau = \frac{Q}{Q_t} \times 273.16$$

where Q is the heat absorbed from (or rejected to) any given reservoir at temperature τ and Q_t is the heat rejected to (or absorbed from) a reservoir, in equilibrium with water at its triple point. We have thus defined the absolute or thermodynamic temperature scale. The unit of temperature on this scale is the kelvin (K).

Now that this temperature scale has been defined, the efficiency of any reversible heat engine may be expressed in terms of it. According to (17-16) and (17-26), the efficiency of any reversible engine operating between two reservoirs at temperatures τ_2 and τ_1 is

$$\eta = 1 - \frac{\tau_2}{\tau_1} \tag{17-27}$$

Finally, then, comparing this with (17-25) for a Carnot engine, we conclude that this temperature scale is the same as that defined in terms of the constant-volume gas thermometer. This then justifies our usage of the kelvin for this unit of temperature.

Example 17-11 What is the maximum efficiency of a steam engine that utilizes steam from a boiler at 480 K and exhausts at 373 K?

Solution The efficiency of this engine is limited by that of a reversible engine that operates between reservoirs at the two temperatures $T_1 = 480$ K and $T_2 = 373$ K. Hence, by use of (17-25), we obtain for this maximum efficiency

$$\eta = 1 - \frac{T_2}{T_1} = 1 - \frac{373}{480} = 0.22$$

In other words, no steam engine, or any other engine for that matter, can operate between these two temperature reservoirs with an efficiency higher than 22 percent. Typical efficiencies realized for steam engines are of the order of 15 percent.

Example 17-12 A reversible engine operates between two heat reservoirs at temperatures of 500 K and 300 K and absorbs during each cycle 600 joules of heat at the higher temperature.

(a) What is its efficiency?
(b) How much heat is rejected per cycle at the lower temperature?
(c) How much work is carried out in each cycle?

Solution

(a) Making use of (17-25), we find that

$$\eta = 1 - \frac{T_2}{T_1} = 1 - \frac{300}{500} = 0.4$$

(b) Solving (17-24) for Q_2 we obtain

$$Q_2 = Q_1 \frac{T_2}{T_1} = 600\ \mathrm{J} \times \frac{300\ \mathrm{K}}{500\ \mathrm{K}}$$
$$= 360\ \mathrm{J}$$

(c) The work W per cycle is the difference $(Q_1 - Q_2)$. Hence

$$W = Q_1 - Q_2 = 600\ \mathrm{J} - 360\ \mathrm{J} = 240\ \mathrm{J}$$

Example 17-13 Derive (17-24) for a Carnot cycle.

Solution Referring to the Carnot cycle as represented in Figure 17-11, we see that, during the initial stage $a \to b$, while the gas expands isothermally its energy does not change, and thus, according to the first law, the heat Q_1 absorbed from the T_1-reservoir is the same as the work carried out on it. Hence, by use of (15-24), we have

$$Q_1 = RT_1 \ln \frac{V_b}{V_a}$$

where it has been assumed that only 1 mole of gas is involved. Similarly, the heat Q_2 rejected during the third stage $c \to d$ is

$$Q_2 = RT_2 \ln \frac{V_c}{V_d}$$

Moreover, since along the path $b \to c$ and $d \to a$ the processes are adiabatic,

$$V_b T_1^{3/2} = V_c T_2^{3/2}$$
$$V_a T_1^{3/2} = V_d T_2^{3/2}$$

and by dividing the second of these into the first we obtain

$$\frac{V_b}{V_a} = \frac{V_c}{V_d}$$

Finally, solving these relations, for the ratio Q_2/Q_1, we obtain

$$\frac{Q_2}{Q_1} = \frac{RT_2 \ln (V_c/V_d)}{RT_1 \ln (V_b/V_a)} = \frac{T_2}{T_1}$$

since the ratio of the logarithms of the volumes cancel out. This then establishes the validity of (17-24).

17-11 Summary of important formulas

If an infinitesimal amount of heat is added to a thermodynamic system initially in equilibrium at temperature T, then its entropy change ΔS when equilibrium is reestablished is

$$\Delta S = \frac{\Delta Q}{T} \tag{17-2}$$

If two originally isolated systems A and B are placed into thermal contact so they can exchange heat and/or work, then when they are in equilibrium again the net changes ΔS_A and ΔS_B in their respective entropies satisfy

$$\Delta S_A + \Delta S_B \geqslant 0 \tag{17-6}$$

where the equality sign prevails only if the processes involved are reversible. For an unnatural process such as the flow of heat from a colder to a warmer body, the sum $(\Delta S_A + \Delta S_B)$ would be negative.

The efficiency η of a cyclic heat engine is

$$\eta = \frac{W}{Q_1} = 1 - \frac{Q_2}{Q_1} \tag{17-16}$$

where Q_1, Q_2, and W are defined in Figure 17-8 and represent, respectively, the heat absorbed per cycle at the higher-temperature reservoir, the heat rejected per cycle at the lower-temperature reservoir, and the work carried out per cycle. The most efficient engine is a reversible, or Carnot, engine, and its efficiency is

$$\eta = 1 - \frac{T_2}{T_1} \tag{17-25}$$

where T_1 and T_2 are the respective temperatures of the hotter and cooler heat reservoirs.

QUESTIONS

1. Define or describe briefly what is meant by the following: (a) natural process; (b) reversible process; (c) entropy; (d) Carnot cycle; and (e) Gibbs paradox.
2. Give three examples of irreversible processes; include at least one unnatural process.
3. Imagine taking a motion picture of some particular natural process. If you view this film when it is run backward, you will observe the same natural process but with its sense of time reversed. Is the time-reversed process always natural or can it be unnatural?
4. Is it possible to convert a certain amount of mechanical work entirely into heat? Is the converse—that is, the conversion of a given amount of heat entirely into work—possible? Illustrate your answer by reference to a particular case.
5. State the theoretical and/or experimental basis for the fact that the entropy of any given system is a state function. Discuss briefly how we can make use of this fact in measuring the entropy of certain thermodynamic systems.
6. In our discussion of entropy we assumed that it was an additive function, so that, for example, the entropy of a system consisting of two components is the sum of the entropies of each component. Explain in physical terms why the entropy of a system must be additive.
7. Consider an ideal gas characterized by the parameters N, V, and E. If we divide the gas in half in some arbitrary way, explain why each half must

be characterized by the values $N/2$, $V/2$, $E/2$?

8. Explain why it is that (17-9) is valid for all liquids and gases and is not restricted to the ideal gas.
9. In Example 17-7 it was found that if 1 mole of an ideal gas undergoes a free expansion from an initial volume V_0 to a final volume $2V_0$, its entropy increases by the amount $R \ln 2$. Explain how this is possible in light of the definition of entropy in (17-2) and the fact that no heat is added to the gas in this process.
10. Two gases—identical in all respects and each at the same temperature—occupy the two halves of an insulated container. If the partition between them is removed, what happens to the entropy of the system? Explain.
11. Consider again the physical situation described in Question 10, but suppose this time that the gases are composed of different molecular species. Explain, in physical terms, why it is that when the partition is removed, this time the total entropy of the system increases. How can you reconcile this increase of entropy with the definition in (17-2) and the fact that no heat enters the system from the outside?
12. State the second law of thermodynamics. Explain why it is that certain physical phenomena that are consistent with the laws of mechanics are nevertheless forbidden by this law.
13. If the formula for the entropy S of a monatomic ideal gas in (17-7) is differentiated with respect to E or with respect to V, a positive result is always obtained. This means that the entropy is an increasing function of both the energy and the volume of the gas. Explain in physical terms why this property of the entropy must be true for fluids in general and is not restricted to the ideal gas.
14. In what sense can the human body be considered a thermodynamic engine? Between what two heat reservoirs does it operate? What is the source of the heat we use to perform mechanical work?
15. Show that on a T-S diagram a Carnot cycle must be a rectangle. What is the physical significance, if any, of the area of this rectangle?
16. What happens to the efficiency of a reversible heat engine if the temperature of the reservoir, to which it rejects heat, drops? Assume that the temperature of the other reservoir is fixed.
17. Consider a reversible cyclic engine operating between two heat reservoirs. What is the entropy change of the engine per cycle? What is the sign of the entropy change of the reservoir at the higher temperature?
18. For the case that both engines in Figure 17-10 are reversible, is it possible that the process shown there might be irreversible nevertheless? Explain.
19. Describe in physical terms the relation between the entropy of a system and its "state of order." Does the melting of a piece of ice correspond to an increase or a decrease in order? Explain.
20. Suppose that a real gas is taken *reversibly* around the Carnot cycle, just as in Figure 17-11, but this time with the isotherms and adiabats not corresponding to those of an ideal gas. Why is the efficiency of the associated Carnot engine still given by (17-25), with T_1 and T_2 the temperatures of the two reservoirs?
21. Describe the relation between thermal pollution and the second law. Is it possible, in an industrial process involving heat engines, to do away entirely with thermal pollution? Explain.

PROBLEMS

1. Using the fact that the heat of fusion of ice is 80 cal/g, calculate the change in entropy of 1 kg of water at 0°C, which freezes to form ice at the same temperature. What is the entropy change associated with this system for the reverse process?
2. In a certain process, 50 cal of heat is extracted from a heat reservoir at a temperature of 200°C. What is the entropy change of the reservoir? What is the minimum entropy change of the *remaining* subsystems that participate in this process?
3. Calculate the entropy change associated with the conduction of 1000 cal of heat along a strip of metal from one heat reservoir at 400 K to a second one of 275 K. Assume that the state of the metal is not altered in the process.
4. Show that if μ moles of a substance are heated reversibly at constant volume from an initial temperature T_i to a final one T_f, then, assuming that C_V is independent of temperature, the entropy change ΔS of the system is

$$\Delta S = \mu C_V \ln\left(\frac{T_f}{T_i}\right)$$

5. **(a)** Calculate the change in entropy of 300 grams of tea, which is originally at a temperature of 100°C and cools off to 30°C. Assume $C_V = 1$ cal/g-K. **(b)** What is the change in entropy of the air in the room?
6. If 2 grams of water at 0°C spontaneously freezes to form ice at the same temperature, how high would the resultant ice cube rise if all of the released energy were converted to motion against gravity?
7. Suppose 30 grams of water at 0°C are poured into a cup containing 300 grams of water at an original temperature of 95°C. Assuming no heat losses and that the specific heat of water has the value 1 cal/g-K calculate **(a)** the final temperature of the mixture and **(b)** the change in entropy of all the water.
8. A 400-gram lead slug has a temperature of 80°C when it is dropped into 50 grams of water originally at 100°C. Assuming that C_V for lead has the value $3R$, calculate **(a)** the final temperature of the system; **(b)** the change in entropy of the water; and **(c)** the total change in entropy of the system. Is your answer consistent with the second law?
9. Making use of the values for heat of vaporization in Table 16-5, calculate the changes in entropy associated with the boiling of 1 mole of each of the following substances. Assume in each case that the liquid and vapor are both at the boiling point corresponding to a pressure of 1 atm: **(a)** O_2; **(b)** N_2; and **(c)** Ag.
10. Repeat Problem 9, but this time calculate the change in entropy associated with the melting of 1 kg of each of the three substances listed. Assume that the values for the heats of fusion in Table 16-5 are applicable.
11. Show that the energy E of a monatomic ideal gas confined to a volume V and consisting of N atoms may be expressed in terms of its entropy S by

$$E = N^{5/3} V^{-2/3} \exp\left[\frac{2S}{3Nk}\right]$$

12. One mole of ^{4}He at an original temperature of 0°C and confined to a volume of 50 liters is compressed reversibly and isobarically at its original pressure to a final volume of

30 liters. (a) What is the gas pressure? (b) What is the final temperature? (c) What is the change in the entropy of the gas?

13. Suppose that 1 mole of a monatomic ideal gas is compressed isothermally and reversibly at a temperature of 50°C from initial volume V_0 to the final volume αV_0, where $0<\alpha<1$. (a) What is the change in entropy of the gas in this process? (b) What is the change in entropy of the heat reservoir?

14. Consider, in Figure 17-13, an imaginary partitioning of a box of volume V and containing N atoms of a monatomic ideal gas of total energy E into two parts of respective volumes αV and $(1-\alpha)V$ (where $0<\alpha<1$).

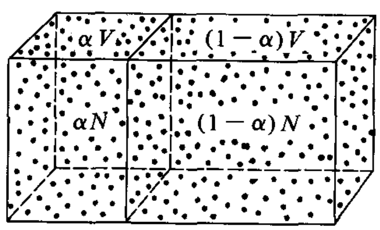


Figure 17-13

(a) Assuming that the constant C in (17-10) depends on the number of atoms N, show, by using the fact that the total entropy of the system must be the sum of the entropies of its two components, that

$$C(N)-C(\alpha N)-C((1-\alpha)N)$$
$$=\tfrac{5}{2}\alpha Nk\ln\alpha+\tfrac{5}{2}(1-\alpha)Nk\ln(1-\alpha)$$

for all values of α satisfying $0<\alpha<1$.

(b) Show that the choice

$$C(N)=-\frac{5}{2}Nk\ln N$$

satisfies this condition and thus confirm the validity of (17-7).

15. Consider again the situations shown in Figure 17-13, but suppose this time that the partition is real. Assume that the temperatures of the gas on the two sides are the same.
   (a) What change in entropy is associated with removal of the partition if identical molecules are on both sides of the container?
   (b) Show that the change in entropy ΔS of the total system upon the removal of the partition if different gases occupy the original two compartments is

$$\Delta S=-\alpha Nk\ln\alpha$$
$$-(1-\alpha)Nk\ln(1-\alpha)$$

   and show that this is positive for all α in the range $0<\alpha<1$.

16. Repeat Problem 15, but this time assume that prior to the removal of the partition the temperature of the gas on the left-hand side has the value T_1 and that on the right has the value T_2.

17. One mole of ^{4}He and one mole of ^{3}He occupy the neighboring compartments of an insulated container separated by an insulating partition. Assume that they both have the same initial volume V_0 and that the initial temperature of ^{4}He is 0°C while that of the ^{3}He is 50°C.
   (a) Calculate the final temperature of the gas if the partition is removed.
   (b) What is the change in entropy of the total gas associated with the removal of the partition?

18. Show that the entropy of a diatomic ideal gas is

$$S=Nk\ln[E^{5/2}V/N^{7/2}]$$

by use of the fact that its energy is $E=5NkT/2$. (*Hint*: Follow the same arguments as used to derive (17-7), but with the altered formula for the energy.)

19. Making use of the result of Problem 18, show that the entropy change ΔS

as given in (b) of Problem 15 is still applicable if the two gases are diatomic.

***20.** Consider two gases confined to the compartments of an insulated container and separated by an insulating, but freely movable, partition. Assuming that the temperatures of the gases have the same value T_0 but that the pressure of one of the gases is originally slightly higher than that of the other, prove by use of the second law that the direction of motion of the piston is such as to equalize the pressure. Assume, if you wish, that both gases are ideal.

***21.** Consider again the physical setup of Problem 20, but this time assume that the partition is diathermal and rigidly fixed. If the temperature of one of the gases is initially slightly above that of the other, prove by use of the second law that the flow of energy is from the hotter to the colder gas.

22. Consider identical ideal gases, each of N molecules, occupying the two halves of a container separated by a diathermal partition. If T_1 and T_2 are the initial temperatures of the two gases and if T_0 is the temperature of one of them finally:

(a) Show by use of energy conservation that the temperature of the other is $[(T_1+T_2)-T_0]$.

(b) Calculate the change in entropy ΔS as a function of T_0.

(c) Show that ΔS is maximum for

$$T_0 = \frac{1}{2}(T_1 + T_2)$$

23. Calculate the efficiency of a cyclic heat engine that during each cycle absorbs 800 joules of heat at the high-temperature reservoir and carries out 150 joules of mechanical work. How much heat is rejected at the lower-temperature reservoir?

24. In a certain household refrigerator, suppose that the coils are at a temperature of 20°F and the compressed gas is at 80°F. If 200 cal of heat are to be taken from the low-temperature reservoir per cycle, what is the minimum amount of heat that must be rejected per cycle at the higher-temperature reservoir?

25. A reversible engine runs between two heat reservoirs at the respective temperatures of 20°C and 80°C. If 10^4 cal are rejected to the cold reservoir during each cycle calculate:

(a) The efficiency of this engine.

(b) The heat absorbed at the hot reservoir and the work carried out per cycle.

26. How much work must be carried out per cycle if 90 cal of heat are transferred from a cold reservoir at -10°C to one at 60°C by use of a reversible refrigerator?

27. Calculate the *minimum* amount of work required to extract 100 cal of heat per cycle from a body at -20°C if the temperature of the environment is 40°C. Neglect the change in temperature of the body.

†28. One mole of ^{4}He is used in a Carnot cycle to drive a reversible engine between two reservoirs at the respective temperatures 500 K and 200 K. If along the upper isotherm (the path *ab* in Figure 17-11), the gas goes from an initial volume of 1.5 liters to 4 liters, calculate **(a)** the heat absorbed from the hot reservoir; **(b)** the heat given up to the cold reservoir; and **(c)** the work carried out per cycle.

29. A reversible engine operates with an efficiency of 50 percent. If during each cycle it rejects 150 cal at a heat reservoir maintained at 30°C:

(a) What is the temperature of the other reservoir?

(b) How much work does it carry out per cycle?

18 Waves

What are the wild waves saying?
J. F. CARPENTER

18-1 Introduction

Under certain conditions, matter in bulk can undergo a very distinctive and highly organized type of motion, for which the term *wave motion* is reserved. In this type of motion the constituent particles in a local region of the *medium* undergo collective and, as a rule, oscillatory motions about their equilibrium positions. The propagation of this motion from point to point in the medium is the important characteristic that defines the wave.

To understand the nature of wave motion more fully, consider a medium of some type that is originally in a state of thermal and mechanical equilibrium; for example, the undisturbed water in a reflecting pool, or the nonvibrating string of a violin, or the stationary air in a room. If a disturbance is created at some point in the medium, the particles in the immediate neighborhood of this point will move about their equilibrium positions and in the process disturb other nearby particles. In turn, the motion of the latter will disturb other particles, and in this way the original disturbance, which may have been highly localized, is propagated through the medium. It is important to note that the constituent particles themselves do not propagate through the medium; only the disturbance does. Hence *wave motion is the propagation of*

motion, or of energy, through a medium without an associated propagation of matter. The medium itself serves only as the vehicle for the transport of energy.

In addition to *mechanical waves* of this type, which involve the propagation of organized motion in a medium, there are a number of other types that are also of importance in physics. One of these, which we shall study in some detail in Chapters 29 and 30, is known as *electromagnetic waves* and this includes visible light, radio waves, and X-rays. For these, no medium is required and, as we shall see, electromagnetic waves are associated with the transport of electrical energy from point to point in *empty space.* A second type is known as *matter waves.* The existence of these waves was first predicted in 1924 on theoretical grounds by Louis-Victor de Broglie (1892–) and they were subsequently observed under controlled laboratory conditions in 1927 by Davisson and Germer and by G. P. Thomson. The field of study concerned with these waves is *quantum mechanics.* More recently, research has produced evidence which appears to indicate the existence of *gravitational waves.* These waves, whose existence was predicted theoretically by Einstein in 1916, seem to originate in the center of our galaxy, the Milky Way, and are associated with the propagation of pure gravitational energy through intergalactic space. The detailed nature of these waves is not well understood, but studies to clarify the matter are currently (1973) under way in many laboratories.

In this chapter we are concerned mainly with describing waves and their more important characteristics. Thus we shall confine ourselves to the conceptually simplest type, namely those which involve a material medium.

18-2 Physical description of waves

To introduce some of the physical ideas necessary for an understanding of wave motion let us first examine qualitatively a number of different types of waves.

Consider first, in Figure 18-1, the familiar sight of shallow waves on the surface of a body of water. Here we see that as a result of falling rain droplets circular waves propagate radially outward from the points of the initial disturbance. Each wave consists of a sequence of concentric circular crests and troughs, which travel radially outward from the point of the disturbance with a definite speed. The fact that no appreciable horizontal motion of water is associated with water waves is easily confirmed by observing the motion of a small floating object, such as a piece of wood or an oil droplet. As the water wave passes, such an object will go vertically up and down as the crest and the associated trough go by, but will undergo very little lateral motion. This observation also confirms the fact that associated with this wave there is a net propagation of energy but not of matter.

As a second case, consider wave motion along a thin, elastic string or wire

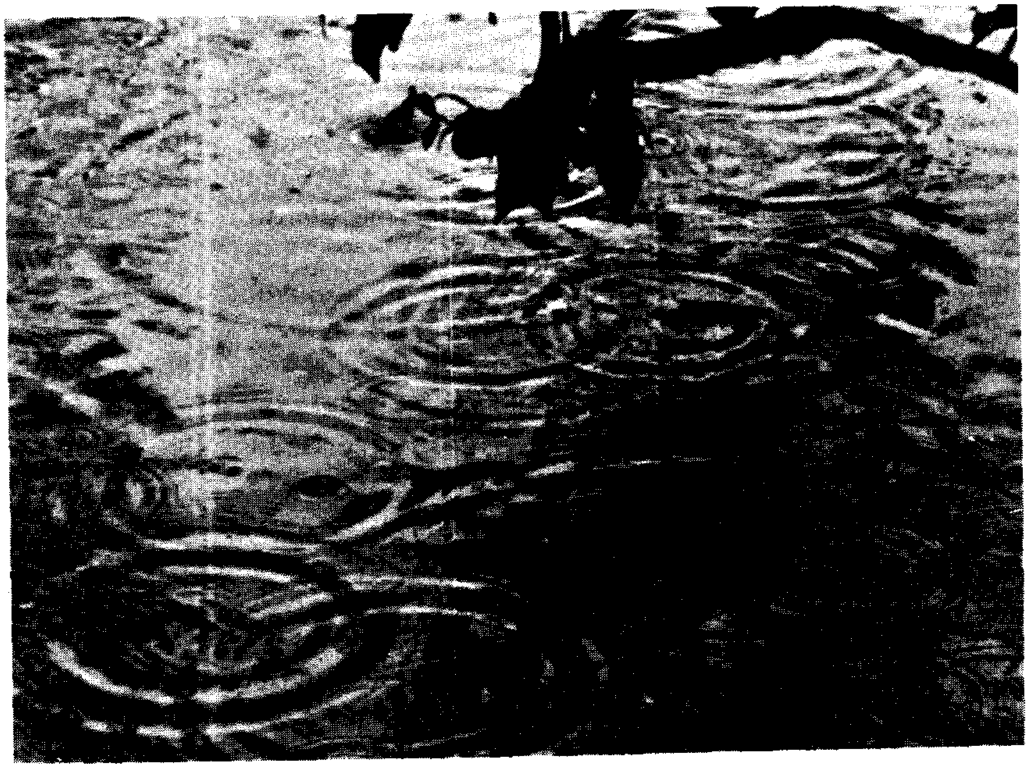


Figure 18-1 *Water waves produced by falling rain droplets.* (*From Project Physics, Holt, Rinehart and Winston, New York.*)

that is suspended between two fixed points and kept under tension. A violin string is an example of this type. Figure 18-2a shows, at a given instant, a *wave pulse*, which is traveling along such a string at a velocity u. As shown in Figure 18-2b, this means that a time interval Δt later the pulse has moved to the right a distance $u\ \Delta t$. The particles that comprise the string do not move along the direction of the wave, but rather they move at right angles to this direction. Just how this up-and-down motion of the individual particles is converted into the horizontal motion of the wave pulse is illustrated in Figure 18-3. The solid curve represents the wave pulse along the string at some instant, and the dashed curve shows the same wave pulse at a slightly later time Δt. As is indicated by the arrows in the figure, for wave motion to the right, on the

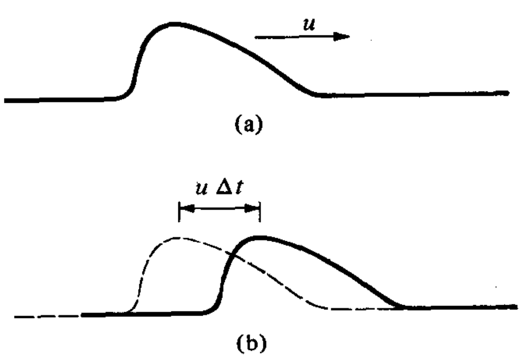


Figure 18-2

Figure 18-3

left-hand side of the wave pulse the individual particles move downward, while on the right-hand side there is a corresponding upward motion. For wave motion to the left, the motion of the particles is in the opposite direction.

A third example of wave motion deals with sound waves in liquids and gases. Consider by way of example, in Figure 18-4a, a long cylindrical pipe filled with a gas, closed at one end and with a movable piston at the other. To produce a sound wave suppose, as shown in Figure 18-4b, that the piston is suddenly plunged into the gas so that the gas density immediately to the right of the piston rises momentarily. The gas in the tube now finds itself instantaneously in a nonequilibrium state, with a higher density in the region near the piston. In an effort to regain thermal equilibrium, this compressed gas expands and in so doing will compress the gas layer immediately to its right, and so forth. In this way, then, a *compressional* pulse—that is, a pulse of gas with a higher than normal density—travels down the tube. Moreover, as shown in Figure 18-4c, as the piston is pulled back to its original position, the gas density in the left part of the tube will be below the equilibrium value. As the air expands to fill this region of lower density, a *rarefaction* pulse—that is, a region of lower than normal density—also propagates along the tube. Thus, by simply plunging in and then withdrawing the piston we can cause an inhomogeneity in density to travel along the tube. A wave of this type, in which a pressure or density inhomogeneity travels through a gas or a liquid, is known as a *sound wave* or as an *acoustical* wave. We hear sounds as a result of the air pressure variation on our eardrums produced by such a wave.

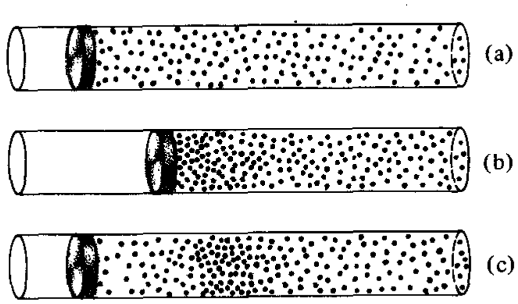


Figure 18-4

It is of interest to note that for acoustical waves the motions of the molecules of the medium are along the direction of motion of the wave. This is in contrast to the waves on a string, where the motion of the constituent particles is at right angles to this direction. For water waves the situation is more complex, and the motions of the constituent particles cannot be characterized as being either along or perpendicular to the direction of propagation of the wave.

18-3 Wave parameters

All wave motion involves the propagation of a disturbance of some type. The *wave velocity* **u** with which such a disturbance travels through a medium is

thus an important parameter of wave motion. To simplify the following discussion of wave velocity it is convenient to assume at first that each point of the wave travels at the same velocity $\mathbf{u}$. We shall say that the medium is *nondispersive* if each element of a wave pulse traveling through the medium does so at the same velocity $\mathbf{u}$. By constrast, when a wave pulse travels through a dispersive medium, different parts of the wave will, in general, travel at different speeds. This means that if a wave pulse travels through a dispersive medium its shape will, in general, change in the course of time. An example of a dispersive medium is a string whose mass density varies along its length. As will be seen below, a wave pulse traveling along this string will not maintain its shape since different parts of the pulse will, in general, travel at different speeds. In the above discussion of the wave pulse on a string in Figure 18-2 it was assumed that the shape of the pulse did not vary in time, and hence that the medium was nondispersive. For reasons of simplicity, in the following unless a statement is made explicitly to the contrary, the underlying medium will always be assumed to be nondispersive.

A second important characterization of a wave has to do with the direction of motion of the particles in the medium relative to that of the wave velocity $\mathbf{u}$. If this motion is at right angles to the direction of $\mathbf{u}$, then the wave is said to be *transverse*. According to the discussion of Section 18-2, the waves on a vibrating string are transverse. Electromagnetic waves are a second example of transverse waves. On the other hand, if the motions of the constituent particles of the medium are along the direction of the wave velocity $\mathbf{u}$, then the wave is said to be *longitudinal*. One example of a longitudinal wave is a sound wave. A second example is shown in Figure 18-5, in which a compressional wave travels along a coiled spring.

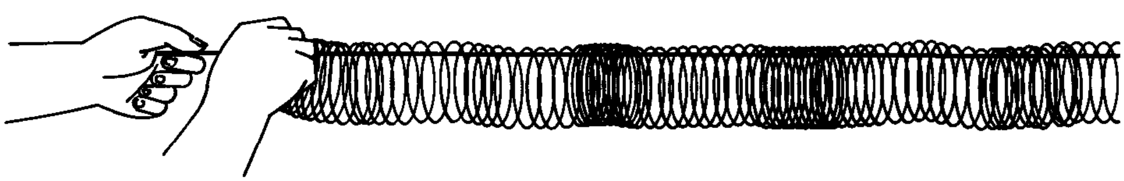


Figure 18-5 *Longitudinal compression waves traveling left to right on a slinky. Note the dispersion of the waves.*

Besides longitudinal and transverse waves, more complicated types of wave motion are also possible. For reasons of simplicity, however, we shall not concern ourselves with these.

In a transverse wave the motion of the constituent particles is at right angles to the wave velocity $\mathbf{u}$. If the motions of all particles involved are in addition parallel to each other, then the wave is said to be *plane-polarized* along the direction of this motion. The plane determined by the wave velocity $\mathbf{u}$ and the direction of motion of the particles is known as the *plane of polarization.* Consider, for example, a wave traveling along the string in Figure 18-6a. Here the plane of the wave pulse lies in the x-y plane and the pulse travels along the positive sense of the x-axis. Hence this particular transverse wave is *plane-polarized in the x-y plane* since the motion of the constituent particles

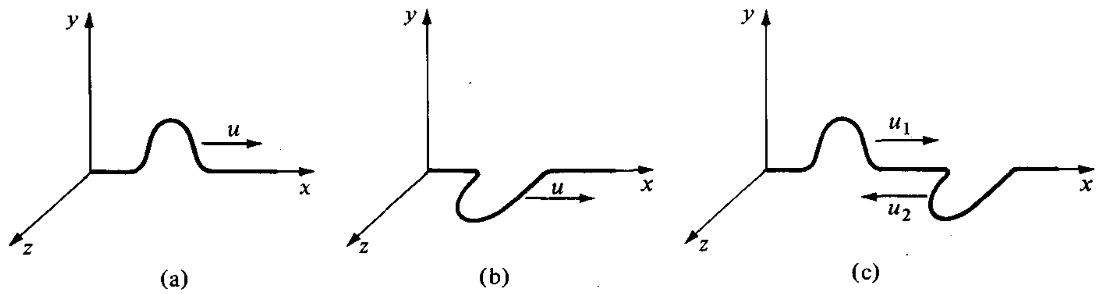


Figure 18-6

are parallel to each other and to the y-axis. Figure 18-6b shows a wave pulse traveling along the positive sense of the x-axis, but this time the motions of the particles are along the z-direction. This wave is plane-polarized in the x-z plane. Finally, Figure 18-6c shows a transverse wave that is *not* plane-polarized since for it some of the particles vibrate along the y-axis and others along the z-axis.

Note that the notion of polarization has been defined only for transverse waves. There is no meaning associated with the state of polarization of a longitudinal wave.

18-4 Wave velocity and the wave equation

Basic to a quantitative description of all wave motion is an appropriate *wave equation*. The purpose of this section is to describe such an equation for the particular case of wave motion along a string. It will be apparent as we go along, however, that the form of this equation is very generally valid and is applicable to other types of wave motion as well.

Consider, in Figure 18-7, a plane-polarized wave pulse traveling at the velocity **u** along an elastic string. Let us set up a coordinate system with the

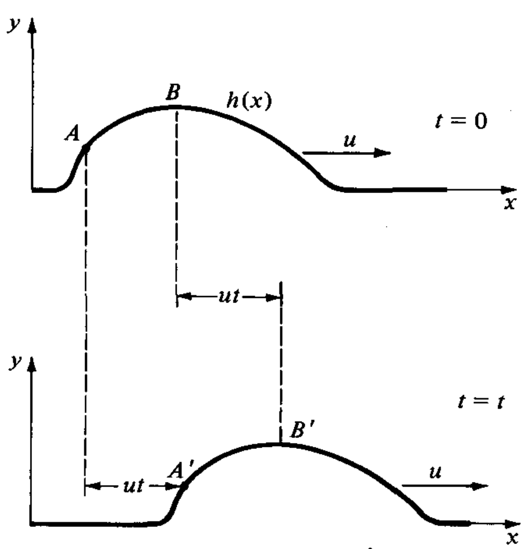


Figure 18-7

x-axis defined by the equilibrium position of the string and so that the wave pulse lies in the x-y plane and travels along the positive sense of the x-axis. To describe this wave in quantitative terms, it is necessary to specify the displacement from equilibrium of each point x of the string at any time t. Since for the coordinate system shown in the figure the displacement from equilibrium of any element of the string is precisely its y-coordinate, it follows that a complete specification of the wave involves a single function $y = y(x, t)$ that gives the displacement of each element of the string at any position x and at any time t. The equation that determines this displacement $y(x, t)$ is known as the *wave equation.*

The problem of deriving a wave equation is in general somewhat complex and the special case of waves on a string is discussed in Appendix F. It is established there that if Newton's second law is applied to a small element of the string (Figure 18-8), then if the displacement $y(x, t)$ of that element is not too large, its motion is governed by the wave equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{u^2}\frac{\partial^2 y}{\partial t^2} \tag{18-1}$$

with u a constant defined in terms of the tension T_0 under which the string is kept and its mass per unit length μ by

$$u = \sqrt{\frac{T_0}{\mu}} \tag{18-2}$$

The parameter u has the dimensions of velocity and as will be confirmed below it represents physically the velocity of waves along the string. The quantity $\partial^2 y/\partial t^2$ in (18-1) represents the up-and-down acceleration of the element of the string at the position x. The associated velocity $\partial y/\partial t$ of this element is not to be confused with the velocity u of the waves along the string. As illustrated in Figure 18-3, it is the collective effect of the up-and-down motion at the velocity $\partial y/\partial t$ of the various elements of the string that manifests itself as a wave traveling along the string at the velocity u.

Although derived only for the vibrating string, with an appropriate reinterpretation of the displacement $y(x, t)$ the relation in (18-1) is found to be very generally valid for waves in all types of nondispersive media. Thus an

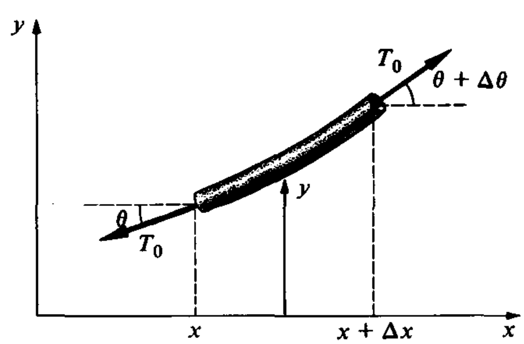


Figure 18-8

analysis of the type in Appendix F, but for water waves in a shallow channel of depth h, shows that (18-1) is again applicable but with the wave velocity

$$u = \sqrt{gh} \tag{18-3}$$

Similarly, for sound waves traveling through a dilute gas, the wave velocity u is found to be

$$u = \left[\frac{\gamma kT}{m}\right]^{1/2} \tag{18-4}$$

where k is Boltzmann's constant (1.38×10^{-23} J/K), T is the absolute temperature, m is the average molecular mass, and γ is the ratio of specific heats as given in Table 16-2. Table 18-1 lists the speed of sound waves in several gases and liquids.

Table 18-1 Velocity of sound (at 0°C and atmospheric pressure)*

Substance	*Density* (kg/m³)	*Velocity* (10^3 m/s)
Air	1.29	0.331
Argon	1.78	0.308
Chlorine	3.21	0.206
Hydrogen	0.09	1.28
Neon	0.90	0.435
Mercury	1.35×10^4	1.45
Water	1.0×10^3	1.43

*Adapted from *Handbook of Chemistry and Physics*, 50th ed. Cleveland: Chemical Rubber Publishing Company, 1970.

In order to get a better feeling for the wave equation in (18-1), suppose as shown in Figure 18-7 that at $t = 0$ the wave pulse on a string has a particular form $h(x)$. In terms of the displacement $y(x, t)$ this means that $y(x, 0) = h(x)$, with $h(x)$ a given function. Since the medium is presumed to be nondispersive, this pulse will travel along the string, without alteration in shape, at the velocity u in (18-2). Hence, the displacement $y(x, t)$ at any time t must be the translation of the original pulse a distance ut along the direction of travel of the wave. As shown in Figure 18-7, for wave motion to the right, any two points such as A and B on the original wave pulse will go into corresponding points A' and B' with the same y-coordinates, but translated a distance ut to the right. In mathematical terms this means that regardless of the shape of the wave pulse, at time t, $y(x, t)$ is given by

$$y(x, t) = h(x - ut) \tag{18-5}$$

The fact that the wave pulse in Figure 18-7 was assumed to be traveling along the positive sense of the x-axis is, of course, not essential. A similar argument applies if the wave pulse travels in the opposite direction. The result

for this case is

$$y = y(x, t) = h(x + ut) \tag{18-6}$$

Figure 18-9 shows what would actually happen if a wave pulse is initiated in a wire by displacing a portion of that wire (Figure 18-9a) and releasing it at rest. If $h(x)$ is the initial pulse, then at any subsequent instant the displacement $y(x, t)$ is the combination $\frac{1}{2}[h(x - ut) + h(x + ut)]$. As shown in Figure 18-9b, this represents two waves, each half the size of the original wave, one traveling to the left and the other to the right (see Example 18-4).

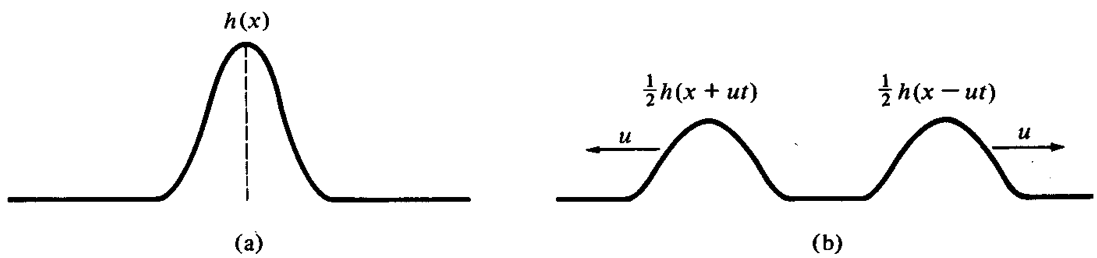


Figure 18-9

An important feature of the displacements $y(x, t)$ in (18-5) and (18-6) is that the variable which characterizes the space-time variation of these waves is either $(x - ut)$ or $(x + ut)$; no other space-time variation occurs. Thus, for y_0 and α constants, the wave amplitudes $y = y_0 \sin[(x - ut)\alpha]$ and $y = y_0 \times \cos[(x + ut)\alpha]$ represent possible waves, whereas the amplitude $y = y_0 \times \sin[(x^2 + u^2y^2)\alpha]$ does not. It is established in Example 18-5 that any (differentiable) function of either of these space-time variables satisfies (18-1). Hence, regardless of the form of the initial wave pulse $h(x)$, the quantities $h(x - ut)$ and $h(x + ut)$ satisfy the wave equation in (18-1). The former represents a wave traveling along the positive sense of the x-axis and the latter represents one traveling to the left.

Example 18-1 Calculate the velocity of waves in:
(a) A region of a lake in which the depth of the water is 2.0 meters.
(b) Air at 30°C.

Solution
(a) Substitution into (18-3) leads to

$$u = \sqrt{(9.8 \text{ m/s}^2) \times 2.0 \text{ m}} = 4.4 \text{ m/s}$$

(b) To determine the velocity of the sound waves in air we make use of (18-4). Using the value $\gamma = 1.4$, and assuming that the average atomic weight of air molecules is 29, we find that

$$u = \left[\frac{\gamma kT}{m}\right]^{1/2} = \left[\frac{1.4 \times (1.38 \times 10^{-23} \text{ J/K}) \times (303 \text{ K})}{29 \times 1.67 \times 10^{-27} \text{ kg}}\right]^{1/2} = 348 \text{ m/s}$$

The fact that this is somewhat larger than the value listed in Table 18-1 is due to the difference in temperature.

Example 18-2 A thin, copper wire of length 10 meters and of mass 0.8 kg is suspended between two fixed points and maintained under tension T_0. If it takes 0.5 second for a small wave pulse that is started at one end to travel to the other, calculate the tension T_0 in the wire.

Solution Since it takes the wave pulse 0.5 second to travel a distance of 10 meters, the velocity u of a wave on this wire is

$$u = \frac{10\ \text{m}}{0.5\ \text{s}} = 20\ \text{m/s}$$

Also, it follows from the given data that the mass per unit length μ of the wire is

$$\mu = \frac{0.8\ \text{kg}}{10\ \text{m}} = 0.08\ \text{kg/m}$$

Solving (18-2) for T_0 and substituting these values we find that

$$T_0 = \mu u^2 = (0.08\ \text{kg/m}) \times (20\ \text{m/s})^2 = 32\ \text{N}$$

Example 18-3 Suppose that the pressure $p(x, t)$ in a long, narrow tube, such as in Figure 18-4, varies with position x measured along the tube and with time t in accordance with the formula

$$p = p_0[1 + \epsilon \cos(kx - \omega t)]$$

where p_0 is the equilibrium pressure in the tube, ϵ is a small, dimensionless constant, and k and ω are parameters having the numerical values $k = 20/\text{m}$, and $\omega = 6.0 \times 10^3/\text{s}$. Assume that this represents a sound wave traveling along the tube.

(a) Find the velocity of the wave.

(b) Show that p satisfies the wave equation.

Solution

(a) Since $p = p(x, t)$ is assumed to represent a wave, it can depend only on the argument $(x - ut)$ or $(x + ut)$. Examination of the given form for p shows that its space-time dependence is $(kx - \omega t)$ and that this may be written in the equivalent form $k(x - \omega t/k)$. Hence, comparison with (18-5) shows that the wave velocity u is

$$u = \frac{\omega}{k} = \frac{6 \times 10^3/\text{s}}{20/\text{m}} = 300\ \text{m/s}$$

(b) Making use of the formulas for differentiating the trigonometric functions in (6-17) and (6-18), and recalling that in taking the partial derivative with respect to one variable the other must be kept constant, we obtain

$$\frac{\partial p}{\partial x} = -p_0\epsilon k \sin(kx - \omega t) \qquad \frac{\partial^2 p}{\partial x^2} = -p_0\epsilon k^2 \cos(kx - \omega t)$$

and

$$\frac{\partial p}{\partial t} = p_0\epsilon\omega \sin(kx - \omega t) \qquad \frac{\partial^2 p}{\partial t^2} = -p_0\epsilon\omega^2 \cos(kx - \omega t)$$

Hence, since $u = \omega/k$, it follows that the given form for p satisfies the wave equation

$$\frac{\partial^2 p}{\partial t^2} = u^2 \frac{\partial^2 p}{\partial x^2}$$

Example 18-4 Show that the wave equation in (18-1) satisfies the *superposition principle* in that if y_1 and y_2 separately satisfy (18-1) and correspond to a wave pulse traveling along a string, then so does their sum, $y = y_1 + y_2$.

Solution Substituting the assumed form $y = y_1 + y_2$ into (18-1), we obtain

$$\frac{\partial^2 y_1}{\partial x^2} + \frac{\partial^2 y_2}{\partial x^2} = \frac{1}{u^2}\frac{\partial^2 y_1}{\partial t^2} + \frac{1}{u^2}\frac{\partial^2 y_2}{\partial t^2}$$

Since both y_1 and y_2 satisfy (18-1) separately, this is an identity and the desired result is established.

More generally, we can show by use of an argument of this type that any linear combination of solutions of the wave equation is itself a solution. Physically this means that if two waves, plane-polarized in the same direction, travel along a string, then the displacement of the resulting wave at any point may be obtained by adding together the displacements of the individual waves. We shall discuss this matter further in Section 18-6.

Example 18-5 Show explicitly that if $h(x)$ is any (differentiable) function, then the displacement $y(x, t)$ in (18-5) satisfies the wave equation in (18-1).

Solution Since $h(x - ut)$ is a function only of the variable $(x - ut)$, it follows by use of the rules for differentiation that

$$\frac{\partial}{\partial x}[h(x-ut)] = \frac{\partial}{\partial(-ut)}[h(x-ut)] = -\frac{1}{u}\frac{\partial}{\partial t}[h(x-ut)] \qquad \textbf{(18-7)}$$

where the second equality follows since u is constant. Differentiating this relation with respect to x, we find that

$$\begin{aligned}\frac{\partial^2 h}{\partial x^2} &= \frac{\partial}{\partial x}\left(-\frac{1}{u}\frac{\partial h}{\partial t}\right) = -\frac{1}{u}\frac{\partial}{\partial t}\left(\frac{\partial h}{\partial x}\right) = -\frac{1}{u}\frac{\partial}{\partial t}\left(-\frac{1}{u}\frac{\partial h}{\partial t}\right)\\ &= \frac{1}{u^2}\frac{\partial^2 h}{\partial t^2}\end{aligned}$$

where the third equality follows by use of (18-7). Comparison with (18-1) thus shows that $h(x - ut)$ satisfies the wave equation.

In a similar way it can be shown that any differentiable function of the variable $(x + ut)$ is also a solution of (18-1).

18-5 Sinusoidal waves

A wave of the type shown in Figure 18-2, which corresponds to a propagating disturbance confined to a local region, is known as a *wave pulse.* As the pulse goes by, a given particle of the string goes from a state of rest to one of motion and eventually returns to a state of rest. On the other hand, if there are a number of wave pulses traveling along the string, one following the other, then we speak of a *wave train* propagating along the wire.

A particularly important manifestation of a wave train is called *sinusoidal* and is associated with the choice of a sine or a cosine function for the

displacement $h(x \pm ut)$ in (18-5) or (18-6). A technique for generating such a wave in the laboratory is shown in Figure 18-10. One end of a string, maintained under tension T_0, is attached to the bob of a harmonic oscillator, which is allowed to vibrate up and down with its characteristic angular frequency ω and amplitude A. That particle of the string immediately in contact with the bob will share in this motion and thus will begin to oscillate up and down with the same angular frequency ω and with the amplitude A. In so doing, it will impart this motion to its nearest neighbor, which in turn will impart it to the next particle, and so forth. In this way each element of the string will eventually oscillate with simple harmonic motion of angular frequency ω and amplitude A.

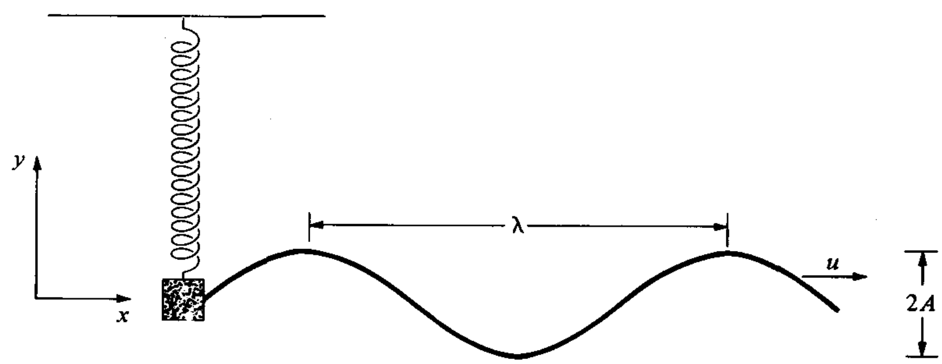


Figure 18-10

To describe this wave in detail we proceed as follows. According to (6-21), the displacement y of a particle undergoing simple harmonic motion at angular frequency ω and with amplitude A may be expressed as $y = A\cos(\alpha - \omega t)$ with α some phase angle. In the present case, the displacement y will also vary with position x along the wire and thus α must be a function of x. To satisfy the general requirement that the space-time variation of a wave can only be in terms of the variables $(x - ut)$ or $(x + ut)$, it follows that α must be proportional to x. Introducing the symbol k for the coefficient of proportionality, we may therefore write

$$y = A\cos(kx - \omega t) \qquad \textbf{(18-8)}$$

The parameter k is known as the *wave number* of the wave. In order that (18-8) actually represent a wave traveling along the positive sense of the x-axis at the velocity u, it is necessary that k have the value

$$k = \frac{\omega}{u} \qquad \textbf{(18-9)}$$

for only then will the wave in (18-8) depend on the single space-time variable, $(x - ut)$. The argument $(kx - \omega t)$ of the cosine function in (18-8) is known as the *phase* of the wave.

A quantity related to the wave number of a sinusoidal wave is the *wavelength* of the wave. We define the wavelength λ of a sinusoidal wave as

the shortest distance between any two points of the wave that have the same displacement and slope at a fixed time t (Figure 18-10). Reference to (18-8) shows then that λ is the shortest positive length for which

$$A \cos (kx - \omega t) = A \cos [k(x + \lambda) - \omega t]$$

Since $\cos (2\pi + \theta) = \cos \theta$ for any θ, it follows that

$$\lambda = \frac{2\pi}{k} \tag{18-10}$$

If we eliminate the wave number k between (18-9) and (18-10), the resulting formula may be expressed in the form

$$\lambda \nu = u \tag{18-11}$$

where ν $(= \omega/2\pi)$ is the frequency of the wave. Its unit is the cycle per second or the hertz (Hz). According to (6-1) and (6-4), ν is also the reciprocal of the period T of the oscillator in Figure 18-10 and thus represents physically the number of times per second that the elements of the string go through a complete cycle. Equivalently, the period $T = 1/\nu$ of the motion also is the period of the wave and thus represents the minimum time that must elapse before the waveform repeats itself everywhere.

Because of (18-10) and (18-11), the sinusoidal wave in (18-8) can be expressed in a number of equivalent ways. Two of these are

$$\begin{aligned} y &= A \cos [k(x - ut)] \\ &= A \cos 2\pi \left[\frac{x}{\lambda} - \frac{t}{T}\right] \end{aligned} \tag{18-12}$$

Most devices used to measure and to detect wave motion are, as a rule, effective only for restricted frequency intervals or equivalently for restricted wavelengths. Thus, our ears are sensitive only to sound waves characterized by frequencies in the range from about 20 Hz to 20,000 Hz. Assuming for sound waves a velocity in air of approximately 300 m/s, the associated wavelength range is from about 15 meters to 1.5 cm. Similarly, our eyes are sensitive to electromagnetic radiation only in the "visible range," corresponding to wavelengths from about 4.0×10^{-7} meter to 7.0×10^{-7} meter. Substituting the known velocity of electromagnetic waves of 3.0×10^{8} m/s into (18-11), we find that this corresponds to a frequency interval from about 4.3×10^{14} Hz to 7.5×10^{14} Hz. Corresponding limitations are applicable for all other devices that we know of that are designed to measure wave motion.

There is a mathematical theorem (due to Fourier) that states that any wave can be expressed as a linear superposition of sinusoidal waves. The importance of sinusoidal waves is due in large part to this result. The idea of thinking of a wave as consisting of a combination of sinusoidal waves is also of practical importance in connection with a study of wave motion in dispersive media. Here we find that wave motion can often be characterized

by assuming that the velocity u of waves in such media is not a constant but depends on wavelength. For example, detailed studies show that the velocity of water waves on the surface of a channel are given by (18-3) only if the depth h of the channel is very small compared to the wavelength λ of the waves. More generally, this velocity is

$$u = \left[\frac{g\lambda}{2\pi}\tanh\frac{2\pi h}{\lambda}\right]^{1/2} \tag{18-13}$$

where tanh x is the hyperbolic tangent of x. Since $\tanh x \cong x$ for small x it is easy to see that if the depth h is very small compared to the wavelength λ, then (18-13) agrees with the previous formula in (18-3). For very short wavelengths ($\lambda \ll h$) the corresponding formula is

$$u = \left[\frac{g\lambda}{2\pi}\right]^{1/2}$$

18-6 Interference

Imagine two waves traveling through the same medium and crossing at some point—for example, the water waves in Figure 18-1. Observation shows that such wave trains go right through each other and that only at the actual points of overlap do they modify one another. Moreover, at each such point of overlap the displacement of the resulting wave is found to be equal to the vector sum of the displacements of the constituent waves at that point. In physical terms this means that when the two waves cross, the displacement that any given particle undergoes is the vector sum of the displacements it would have undergone if each wave had passed separately. This feature that the displacement of two overlapping waves is the sum of the displacements of the individual waves is known as the *superposition principle.* This principle is very generally valid for all wave motion and is not confined to waves in elastic media.

We shall use the term *interference* to refer to the effects produced when two or more waves pass through the same region of space. Interference effects are associated exclusively with wave motion. Indeed, the observation of interference for any physical system implies unambiguously that a wave motion of some type is involved. Thus, the discovery by Davisson and Germer and by G. P. Thompson that an electron beam could exhibit interference by scattering from certain crystals led unequivocally to the conclusion that a wave motion was involved. As noted previously, this was the first experimental indication of the existence of matter waves.

We can demonstrate the consistency of the superposition principle with the wave equation in (18-1) in the following way. Suppose that y_1 and y_2 represent two initially nonoverlapping wave pulses on a string such as those in Figure 18-11a or Figure 18-12a. According to the superposition principle, for times t

that the waves overlap, the resulting amplitude y is given by

$$y = y_1 + y_2 \tag{18-14}$$

Hence, it is necessary that if y_1 and y_2 each satisfy the wave equation in (18-1) then so must their sum. The fact that this is indeed the case has been established in Example 18-4 and is a consequence of the linearity of the wave equation.

By way of example, consider in Figure 18-11a, two identical wave pulses on a string approaching each other. Figure 18-11b shows the situation shortly after they begin to overlap and Figure 18-11c illustrates the situation when the waves overlap in their entirety. At this instant since the displacements y_1 and y_2 of the two waves are everywhere equal, the displacement y of the total wave is twice that of either of the original waves according to (18-14). The waves are said to *interfere constructively* or to undergo *constructive interference* in this case. At the instant of complete constructive interference in Figure 18-11c, the velocity of each element of the string vanishes. Finally, Figure 18-11d shows the situation subsequently when the wave pulses recede from each other and the waves no longer interfere. The situation then is precisely that portrayed in Figure 18-9b.

Figure 18-12 shows the corresponding situation when destructive interference takes place. Part (a) shows two identical pulses with equal but opposite displacements approaching each other, and part (b) shows the situation just as

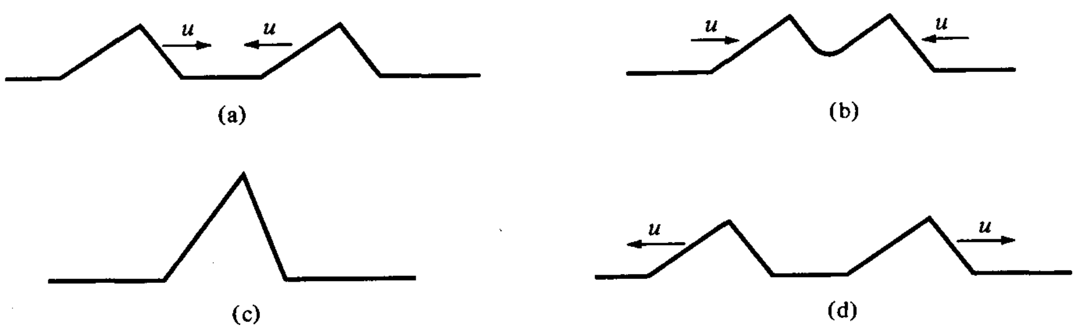


Figure 18-11

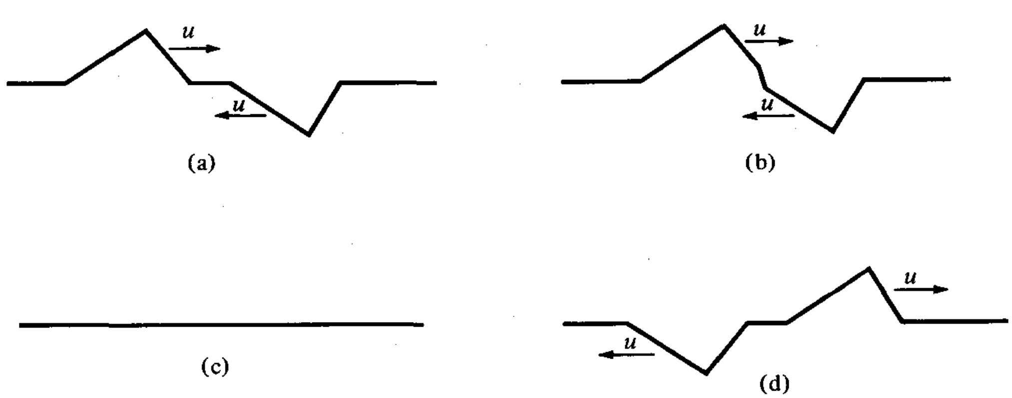


Figure 18-12

they begin to overlap. This time the interference is *destructive.* The situation at the instant of complete overlap is shown in Figure 18-12c. Here the displacement y_1 and y_2 are equal and opposite, and hence the displacement of the wave vanishes. Figure 8-12d shows the situation still later when the two wave pulses have passed through each other and now recede in opposite directions.

It is interesting to view the phenomena of constructive and destructive interference from the viewpoint of energy. As will be illustrated in Section 18-10, in general, both kinetic and potential energy are associated with a wave propagating along a string. The kinetic energy is contained in the up-and-down motions of the various elements of the string (Figure 18-3). Correspondingly, the string has a potential energy by virtue of its being under tension and then stretched as the wave passes by. As a general rule, a given wave pulse will have both kinetic and potential energy. However, if two or more wave pulses overlap, this may not be the case. In particular, at the instant of complete overlap of the two pulses in Figure 18-11c, since the velocities of all elements of the string are zero, the kinetic energy vanishes. Hence at the instant when complete constructive interference takes place, the energy of the wave is all tied up in the elastic energy of the string. On the other hand, for the case of complete destructive interference, at the instant of complete overlap in Figure 18-12c, the string is in its unstretched state so that its potential energy vanishes. In other words, the energy of the wave at the instant of complete destructive interference is all kinetic. Of course, after the waves have passed each other, as in Figures 18-11d and 18-12d, each wave pulse will again have its original potential and kinetic energy.

A phenomenon related to interference is known as *diffraction* and has to do with the ability of waves to bend around sharp edges. The fact that we can hear sounds whose sources are separated from us by an obstruction such as a building or a wall is a manifestation of this phenomenon. We shall study diffraction later in connection with electromagnetic waves.

18-7 Standing waves

Up to this point we have been concerned only with waves propagating in regions of a medium far from its bounding surface. If a wave strikes the bounding surface of a medium, a *reflected* wave is usually observed traveling away from the surface. If the direction of motion of the reflected wave is such that it crosses the incident wave, then interference effects may be observed. The reflected wave depends, in general, on the properties of the medium and that of the surface, and we shall use the term *boundary conditions* when referring to the behavior of the total wave at such a bounding surface. The determination of boundary conditions is, in general, very complex, and in the following only the case of wave motion along a string will be discussed.

Consider, in Figure 18-13a, a wave pulse traveling along a thin wire at an

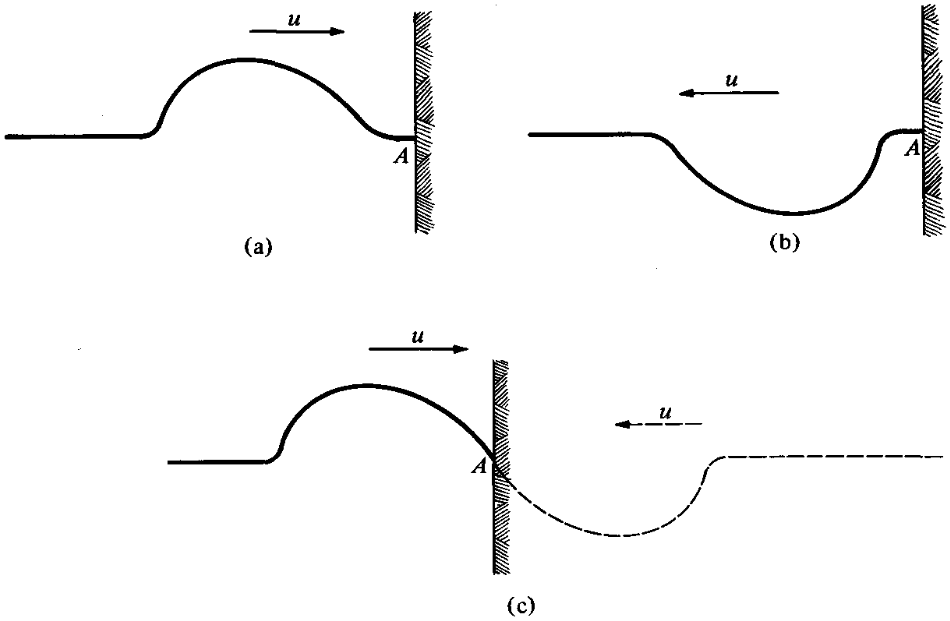


Figure 18-13

instant just before it reaches the fixed endpoint A. The boundary condition here is very simple and is that this endpoint A of the wire is immovable. This means that as the wave reaches the point A, it exerts on it an upward force tending to make the fixed point A follow the upward motion of the wave. However, A is fixed so that this end of the string cannot undergo this motion. By Newton's third law, the contact at A must exert on the string an equal and opposite force, and in this way a reflected wave is set up. Figure 18-13b shows the situation after the complete wave pulse has been reflected at the boundary point A. Note that the reflected pulse is inverted both up-down and forward-backward relative to the incident wave pulse.

An alternate way of viewing this reflection at A can be obtained by use of the superposition principle. For this purpose, think of the wave on the string at all times to be the superposition of two waves traveling in opposite directions. One of these is the actual wave and the other is an imaginary one selected so that the net displacement of the total wave at the point A vanishes at all times t. To achieve this, imagine as in Figure 18-13c the string extended to the right of A and carrying a wave traveling to the left and inverted (up-down and forward-backward) relative to the incident wave and being at all times as far to the right of A as the real wave is to its left. This second wave has been indicated by a dashed line in the figure. In the region to the left of A, then, the superposition of these two waves satisfies the condition that at any time t the displacement of the total wave at the point A vanishes.

Let us now specialize to the case of the sinusoidal wave train $y = A\cos(kx - \omega t)$ in (18-8) traveling along a string fixed at both ends and of length L so that $0 \leq x \leq L$. Because the string is fixed at the endpoints at $x = 0$ and $x = L$, an appropriate reflected wave will be generated, and, based on the above discussion, it must have the form $y = -A\cos(kx + \omega t)$. Note that the

argument of the cosine is now $(kx + \omega t)$, since the reflected wave must travel in a direction opposite to that of the original one. The total wave is known as a *standing wave* and is the algebraic sum of these two:

$$\begin{aligned} y &= y_1 + y_2 \\ &= A\cos(kx - \omega t) - A\cos(kx + \omega t) \end{aligned}$$

Making use of the trigonometric identity

$$\cos(\alpha - \beta) - \cos(\alpha + \beta) = 2\sin\alpha\sin\beta \qquad \textbf{(18-15)}$$

we may reexpress this in the form

$$y = [2A\sin kx]\sin\omega t \qquad \textbf{(18-16)}$$

and this is the formula for a standing wave on a string.

Physically, we can think of a standing wave in the following way. Consider the motion of the particle of the string at a position x. Reference to (18-16) shows that this particle undergoes simple harmonic motion with angular frequency ω and with amplitude $|2A\sin kx|$. In other words, all of the "particles" on the string oscillate up and down with simple harmonic motion of the same angular frequency and phase, but with amplitudes that vary with position. The particles located at the points $x = 0, \pi/k, \ldots$, for example, have zero amplitude and hence do not move at all, while those located at $x = \pi/2k, 3\pi/2k, \ldots$ oscillate with the maximum amplitude of $2A$.

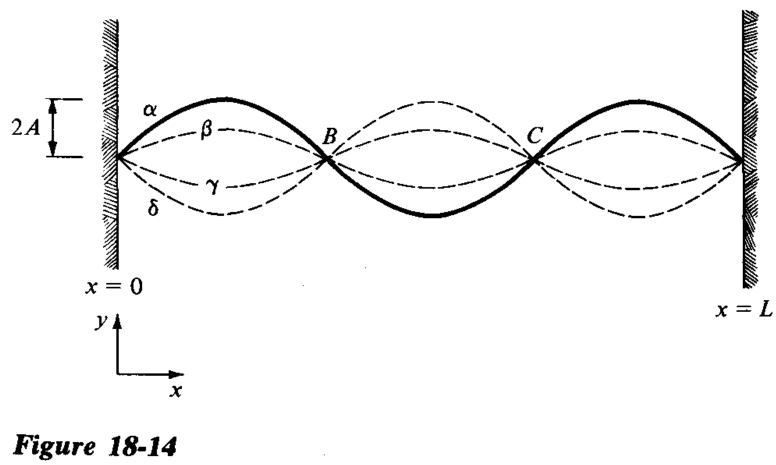


Figure 18-14

Figure 18-14 shows the instantaneous position at four different times of a standing wave on a string. The curve labeled α represents the string at an instant t for which $\omega t = \pi/2$. The curves labeled β, γ, and δ then represent the string at the respective times $\omega t = 3\pi/4$, $5\pi/4$, and $3\pi/2$. Note that the particles at points B and C, for which $x_B = \pi/k$ and $x_C = 2\pi/k$, do not undergo any motion whatsoever. These positions along a vibrating string where no motion occurs are known as *nodes*.

18-8 Normal modes

In order that (18-16) actually represent a standing wave on a string with fixed endpoints at $x = 0$ and at $x = L$, it is necessary that the displacement y vanish at both of these points for all times t. Since the function $\sin kx$ vanishes at $x = 0$, this condition is automatically satisfied at this end. However, to satisfy it at the other, it is necessary to restrict the values which the wave-vector parameter k—and the associated values for ω in (18-9)—can assume. The standing waves associated with any one of these allowable values for k or ω are known as the *normal modes* of vibration of the string.

To determine the normal modes, we make use of the fact that the sine of any integral multiple of π vanishes. Reference to (18-16) then shows that the boundary condition at the right end of the wire at $x = L$ will be satisfied provided that k is restricted to the values

$$kL = \pi, 2\pi, \ldots, n\pi, \ldots \qquad \textbf{(18-17)}$$

Only for these k-values, which determine the normal modes, will (18-16) represent a physically realizable wave on this string.

It is convenient for many purposes to reexpress (18-17) in terms of wavelength and frequency. Making use of (18-10) and (18-11) we find that the wavelengths λ_n $(n = 1, 2, \ldots)$ for standing waves on a string are

$$\lambda_n = \frac{2L}{n} \qquad (n = 1, 2, \ldots) \qquad \textbf{(18-18)}$$

and that the frequencies ν_n associated with these are

$$\nu_n = \frac{u}{\lambda_n} = \frac{u}{2L}n \qquad (n = 1, 2, \ldots) \qquad \textbf{(18-19)}$$

These frequencies in (18-19) are also known as the resonant frequencies of the string. The first of these $\nu_1 = u/2L$ is called the *fundamental*, and the remaining ones $\nu_2 = 2\nu_1$, $\nu_3 = 3\nu_1, \ldots$, are known respectively, as the first, the second, ... *overtone*. The frequencies in (18-19) also form a *harmonic series*, since the ratios of successive frequencies are equal to the ratios of successive integers. The terms *first harmonic, second harmonic*, and so on, are also used for the frequencies $\nu_1, \nu_2, \ldots$.

The physical significance of (18-18) and (18-19) is that only for these wavelengths and associated frequencies can there be standing waves on the string. If, for example, we start a wave on a string, say by plucking it, then the total wave generated must be a linear combination of these *normal modes*. Any other wave would not satisfy the boundary conditions at $x = 0$ and $x = L$, and thus could not exist on the given wire.

Figure 18-15 shows the normal modes associated with the fundamental and the first three overtones of a vibrating string. As defined above, the positions along the string at which no motion takes place are called *nodes*. Thus for

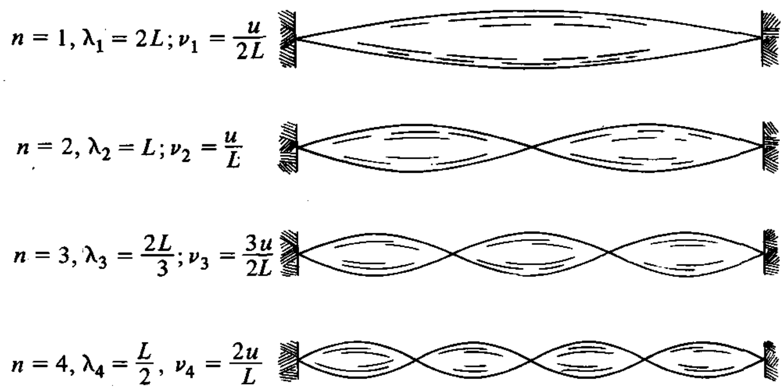


Figure 18-15

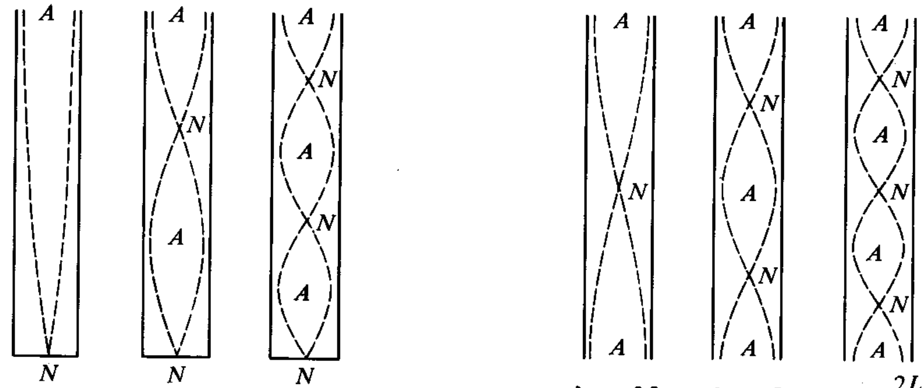
$n = 1$ the standing wave has two nodes one at each end, while for $n = 2$ there are three nodes: one at each end and one in the middle. The positions midway between two nodes, which correspond to maxima in displacement, are called the *antinodes*. In general, the standing wave associated with frequency ν_n has precisely $(n + 1)$ nodes and n antinodes. Since the distance between neighboring nodes or antinodes is exactly half a wavelength, it follows that the standing wave associated with the frequency ν_n corresponds to $n/2$ wavelengths, each of length $2L/n$. As shown in the figure, for example, the standing wave for $n = 4$ has two wavelengths. The standing wave y_n associated with the nth normal mode may be expressed by use of (18-16) through (18-19) by

$$y_n = 2A_n \sin\left(\frac{\pi n}{L} x\right) \sin\left(\frac{\pi u n}{L} t\right) \tag{18-20}$$

where $2A_n$ is the maximum amplitude of this standing wave. Only a linear superposition of waves of this type can exist on a string fixed at both ends.

Standing waves also exist in physical systems other than strings. For example, associated with organ pipes we also find characteristic frequencies and associated overtones. If a compressional wave travels down such a tube, then at a closed end there arises a reflected wave of a nature such that there is no displacement of the air at this end. The situation here is therefore very much analogous to that of a wave reflected at the fixed end of a string. Hence, an organ pipe of length L closed at both ends has standing waves with associated frequencies given precisely by (18-19), only with u now representing the velocity of air in the organ pipe.

Figure 18-16a shows some of the normal modes associated with organ pipes open at one end, and Figure 18-16b shows some of those associated with organ pipes open at both ends. Experiment shows that to a high degree of approximation, at an open end there is an antinode in displacement. Hence, since there must be a node in displacement at the other end, it follows that the fundamental frequency ν_1 for a pipe open at one end must be $u/4L$. As shown in Figure 18-16a, the first overtone is a standing wave with an additional node and antinode, and thus has the frequency $\nu_2 = 3u/4L$. In general, for a pipe



$\lambda_1 = 4L \quad \lambda_2 = \frac{4L}{3} \quad \lambda_3 = \frac{4L}{5}$

$\nu_1 = \frac{u}{4L} \quad \nu_2 = \frac{3u}{4L} \quad \nu_3 = \frac{5u}{4L}$

(a)

$\lambda_1 = 2L \quad \lambda_2 = L \quad \lambda_3 = \frac{2L}{3}$

$\nu_1 = \frac{u}{2L} \quad \nu_2 = \frac{u}{L} \quad \nu_3 = \frac{3u}{2L}$

(b)

Figure 18-16

open only at one end the overtones are odd integral multiples of the fundamental. For the case of a pipe open at both ends it is apparent that the frequencies are the same as those associated with one closed at both ends. The difference is that for the former there is an antinode, instead of a node, at each end. Thus the fundamental frequency is $u/2L$ and the overtones are integral multiples of it.

Example 18-6 Suppose the fundamental frequency of a certain violin string of length 40 cm is 450 Hz. Calculate:

(a) The velocity of waves along the string.
(b) The wavelength λ_1 associated with the fundamental frequency.
(c) The wavelength λ_3 of the third harmonic.

Solution

(a) Solving (18-19) for u, we find for $n = 1$

$$u = 2L\nu_1 = 2 \times 0.4 \text{ m} \times 450 \text{ Hz} = 360 \text{ m/s}$$

since 1 Hz corresponds to one vibration per second.

(b) The wavelength λ_1 associated with this fundamental frequency may be obtained by use of (18-18) with $n = 1$. The result is $\lambda_1 = 2L = 0.8$ meter.

(c) The wavelength associated with the third harmonic may be obtained by a second application of (18-18), this time with $n = 3$. Thus we find that

$$\lambda_3 = \frac{2L}{3} = \frac{2 \times 0.4 \text{ m}}{3} = 0.27 \text{ m}$$

Example 18-7 An organ pipe of length 50 cm is closed at one end and vibrates in its first overtone. Assuming conditions are such that the speed of sound is 330 m/s, calculate the frequency of the tone emanating from the pipe. Calculate also the distance between neighboring nodes under these circumstances.

Solution The situation here is that shown in the middle pipe in Figure 18-16a. Inserting the known values into the given formula for ν_2 we obtain

$$\nu_2 = \frac{3u}{4L} = \frac{3 \times 330\ \text{m/s}}{4 \times 0.5\ \text{m}}$$
$$= 495\ \text{Hz}$$

Reference to the same figure shows that the distance between two nodes is two thirds the length of the pipe, or $(2/3) \times 0.5$ meter $= 0.33$ meter.

18-9 The Doppler effect

It is a familiar experience that the frequency (pitch) of the sound emanating from the whistle of a moving train appears to be higher when it approaches than when it recedes. A similar effect is noted when the listener is moving while the train is at rest. It was first pointed out in 1842 by Christian J. Doppler (1803–1853) that this phenomenon is not restricted to sound waves but is true for wave motion in general. Thus the discovery in 1927 by V. M. Slipher that the observed spectral lines emitted from galaxies are shifted toward lower frequencies, the *red shift,* led to the conclusion that galaxies are generally moving away from each other, and to our present view that the universe is expanding. This effect of a shift in the frequency of a wave as a result of motion of either the source or the observer is known as the *Doppler shift.* In this section we shall discuss this effect only as it relates to sound waves.

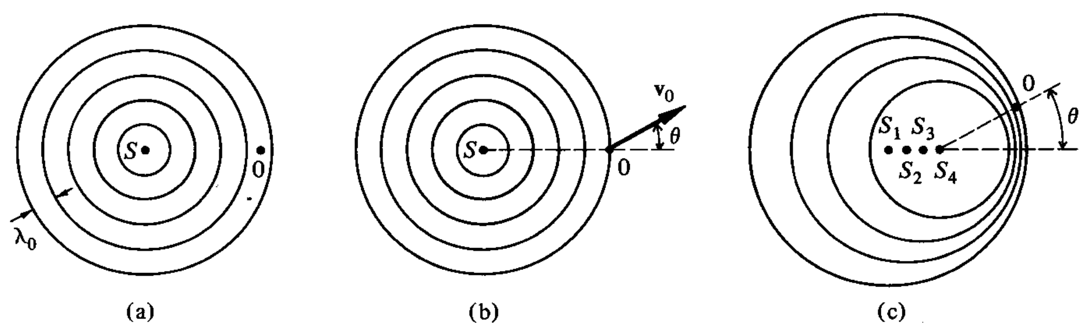


Figure 18-17

Consider first, in Figure 18-17a, a source S of sound of wavelength λ_0 and an observer O, both at rest relative to the medium, air. The series of concentric circles about S represent surfaces of maximum displacement of the sound waves emanating from S, and they are known as *wavefronts.* We can think of each of the circles in the figure as representing a surface of maximum pressure which propagates radially outward from S with the velocity of sound u. The distance between consecutive circles is one wavelength λ_0, and for the given situation this is related to the frequency ν_0 of

the source in accordance with (18-11) by

$$\nu_0 = \frac{u}{\lambda_0} \tag{18-21}$$

Let us now calculate the frequency ν that is heard by the observer O when he travels at a velocity $\mathbf{v}_0$ in a direction making the angle θ with a line drawn from the observer to the source (Figure 18-17b). We assume here that v_0 is less than the speed of sound u in the medium. If the observer were not in motion, then in a time interval Δt, $u\ \Delta t/\lambda_0 = \nu_0\ \Delta t$ waves would pass him by. However, because he has a radial component of velocity $v_0 \cos\theta$ with respect to S, only $(u - v_0 \cos\theta)\ \Delta t/\lambda_0$ actually do. This number, which pass him by per unit time Δt, represents the frequency ν of the sound that he hears. Hence

$$\nu = \frac{(u - v_0 \cos\theta)\Delta t/\lambda_0}{\Delta t} = \frac{u}{\lambda_0}\left(1 - \frac{v_0}{u}\cos\theta\right)$$

which, according to (18-21), may be expressed as

$$\nu = \nu_0\left(1 - \frac{v_0}{u}\cos\theta\right) \tag{18-22}$$

Note that this frequency shift $(\nu - \nu_0)$ ranges from a maximum of $\nu_0(v_0/u)$, when the observer approaches the source ($\theta = 180°$), to a minimum of $-\nu_0(v_0/u)$, when he moves away from it ($\theta = 0$).

The corresponding derivation of the frequency shift associated with the source S in motion relative to the observer is somewhat more complex. As is evident physically and as shown in Figure 18-17c, in this case the circles representing the instantaneous surfaces of maximum pressure are no longer concentric with S. Assuming that the separation distance between the observer and the source is very large compared to a wavelength, the appropriate formula now is

$$\nu = \frac{\nu_0}{1 - (v_s/u)\cos\theta} \tag{18-23}$$

where v_s is the velocity of the source S relative to the medium and θ is the angle defined in the figure. It is assumed here that the velocity v_s of the source is less than the speed of sound u in the medium. Since the derivation of (18-23) is somewhat complex, let us confirm its validity only for the special case $\theta = 0$.

To this end, let us note that during each period $1/\nu_0$ of vibration of the source, it travels a distance v_s/ν_0 toward the listener. In effect, then, the wavelength of the sound wave reaching O is shortened by this distance and has the value $(\lambda_0 - v_s/\nu_0)$. Making use of (18-21), we have then

$$\nu = \frac{u}{\lambda_0 - v_s/\nu_0} = \frac{u/\lambda_0}{1 - v_s/\lambda_0\nu_0} = \frac{\nu_0}{1 - v_s/u}$$

This then confirms the validity of (18-23) for the special case $\theta = 0$.

It is important to note that these Doppler shift formulas in (18-22), when the observer is in motion, and in (18-23), when the source is in motion, are different. This means that there is no principle of "Galilean invariance" here as there is in mechanics. That is, the relative motion between the source and the observer is *not* the only relevant variable; the absolute motion of both source and observer also plays a significant role. In other words, for sound waves in air there is a preferred coordinate system, namely the one at rest with respect to the medium. In view of this, we should not expect—and, indeed, do not find—invariance under Galilean transformations, and hence the difference between (18-22) and (18-23).

It is interesting to note that, nevertheless, if the source and observer approach each other, then regardless of which is in motion there is an increase in frequency. Thus, if the observer approaches the source, then, in (18-22), $\theta = 180°$ and $\nu > \nu_0$. Correspondingly, if the source approaches the observer, then in (18-23) we must take $\theta = 0$ and, again, $\nu > \nu_0$. The physical reason for this is easily seen, since in both cases the distance between succeeding wavefronts is less than λ_0.

It should be emphasized that in the above discussion it has been explicitly assumed that the speeds v_0 and v_s of both the observer and the source are less than is the speed of sound u in the air. If the source travels at a speed greater than the wave speed in the medium, then the surfaces of constant pressure do not have the simple form shown in Figure 18-17. The *sonic boom* associated with jet airplanes traveling at speeds in excess of the speed of sound in air is a consequence of this phenomenon. The nature of the waves emanating from a source in motion at a speed in excess of that of waves in the medium can be seen by reference to Figure 18-18. Here we see the water waves produced by a

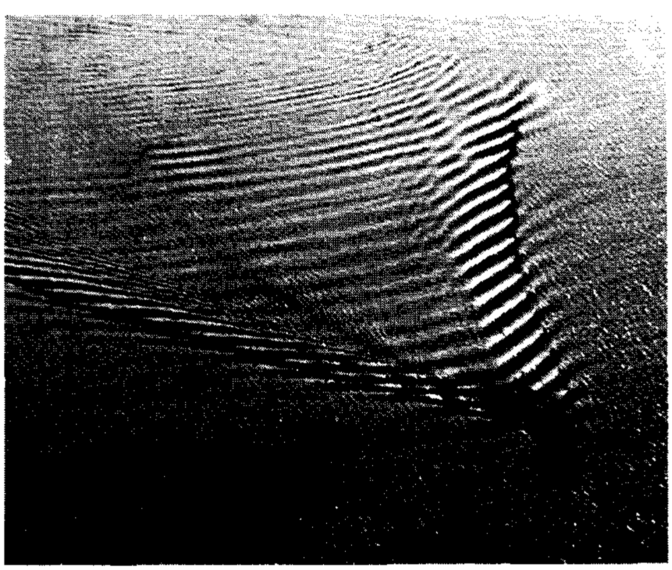


Figure 18-18 *Water waves produced by a boat traveling at a speed greater than that of the waves. (From Project Physics, Holt, Rinehart and Winston, New York.)*

boat traveling over water at a speed greater than the speed of waves in this water. For this case the source, here the boat, travels a greater distance, in any given interval Δt, than does the wave generated at the beginning of that time interval.

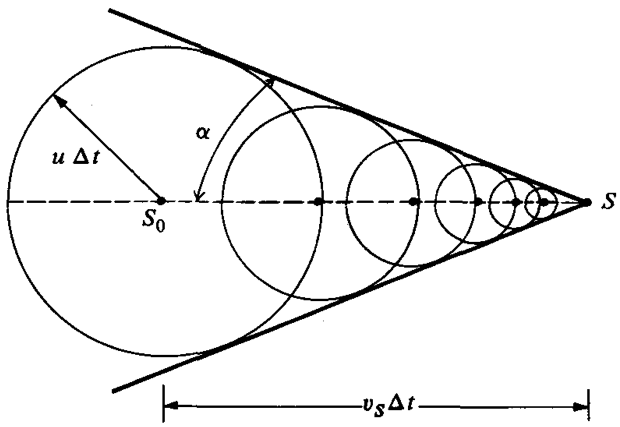


Figure 18-19

To determine the nature of the sound wave for the case $v_s > u$, consider the situation in Figure 18-19. Let S_0 be, at the initial instant, the location of a source that, in a time interval Δt, travels to the final position S at a speed v_s. We assume now that $v_s > u$, with u the speed of sound in air. The circles in the figure show the positions at the end of the time interval Δt of waves generated by the source when it is at various positions along the line S_0S. The envelope of these circles determines the wavefront. This wavefront is evidently a cone whose half angle α may be obtained, by reference to the figure, to be

$$\sin\alpha = \frac{u}{v_s} \tag{18-24}$$

Note that only for $v_s > u$ is there a real angle α associated with this motion. The ratio v_s/u is also known as the *Mach number* and we speak of an airplane traveling at "Mach 2" if its speed v_s is twice that of sound.

†18-10 The energy of a wave

Since wave motion represents the propagation of energy in a medium, the nature of the energy associated with a wave is of considerable interest. To give a feeling for the kinds of physical quantities involved, in this section it will be established that the total energy E associated with small amplitude waves on a string of length L is

$$E = \int_0^L dx\left[\frac{1}{2}\mu\left(\frac{\partial y}{\partial t}\right)^2 + \frac{1}{2}T_0\left(\frac{\partial y}{\partial x}\right)^2\right] \tag{18-25}$$

where μ is the mass per unit length of the string and T_0 its tension. The first

term in the integrand, $\frac{1}{2}\mu(\partial y/\partial t)^2$, represents the kinetic energy per unit length of the string, and the second its potential energy per unit length. The fact that E depends quadratically on the derivatives of the displacement y is a general feature of all wave motion.

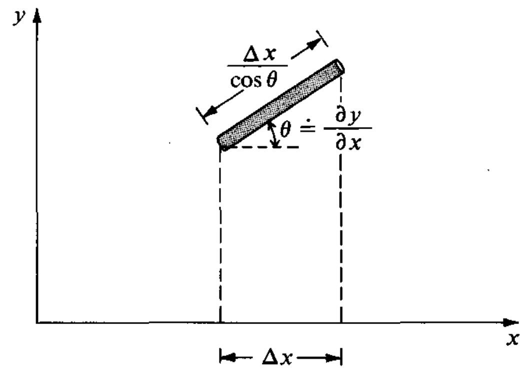


Figure 18-20

To derive this formula, consider in Figure 18-20 an infinitesimal element of length Δx of a string which is located at point x. If at time t it has a vertical displacement y, then its velocity v at that instant is $\partial y/\partial t$. Let us assume in the following that all displacements are sufficiently small so that the angle θ between the given element of the string and the horizontal satisfies the condition $\tan\theta \cong \theta$, and therefore that

$$\theta = \frac{\partial y}{\partial x} \tag{18-26}$$

Now the kinetic energy $\Delta(KE)$ of this element of the string is

$$\Delta(KE) = \frac{1}{2}\mu\,\Delta x\left(\frac{\partial y}{\partial t}\right)^2 \tag{18-27}$$

since its mass is $\mu\,\Delta x$ and its velocity is $(\partial y/\partial t)$. According to (8-12) the potential energy $\Delta(PE)$ of this element of the string is the negative of the work ΔW carried out by the tension T_0 in stretching it from its unstretched length Δx to its stretched length $\Delta x/\cos\theta$. Hence its potential energy is

$$\Delta(PE) = -\Delta W = -T_0\left[\Delta x - \frac{\Delta x}{\cos\theta}\right] = T_0\,\Delta x\left[\frac{1-\cos\theta}{\cos\theta}\right]$$
$$\cong T_0\,\Delta x\left(\frac{1}{2}\theta^2\right)$$

where the last equality follows from the fact that for small θ, $\cos\theta \cong 1-\theta^2/2$. Substituting for θ by use of (18-26), we obtain

$$\Delta(PE) = \frac{1}{2}T_0\,\Delta x\left(\frac{\partial y}{\partial x}\right)^2 \tag{18-28}$$

The result in (18-25) then follows by adding together contributions of the form in (18-27) and (18-28) for each element Δx of the string and taking the limit as the lengths of these elements tend to zero.

It should be noted that the sum of the kinetic energy $\Delta(KE)$ and the potential energy $\Delta(PE)$ of any given element of the string is *not* conserved. Consider a wave pulse traveling along the string. As the pulse passes through any given element, the potential and kinetic energies of that element must increase from zero to positive values, and hence its total energy cannot be conserved. In physical terms this means that energy need not be conserved in a local region of the string. However, as we should expect and as will be confirmed in Problem 37, the total energy of the string is conserved.

Example 18-8 Show that kinetic energy of each element of the string is equal to the potential energy of that element.

Solution Since the space-time variation of the displacement y is exclusively in terms of the variables $(x - ut)$ or $(x + ut)$, it follows that y must satisfy (see Example 18-5)

$$\frac{\partial y}{\partial x} = \pm \frac{1}{u} \frac{\partial y}{\partial t}$$

For either choice of sign then

$$\left(\frac{\partial y}{\partial x}\right)^2 = \frac{1}{u^2}\left(\frac{\partial y}{\partial t}\right)^2 = \frac{\mu}{T_0}\left(\frac{\partial y}{\partial t}\right)^2$$

where the second equality follows by use of (18-2). Comparison with (18-27) and (18-28) then shows that at each point the kinetic energy associated with the wave is numerically equal to its potential energy.

18-11 Summary of important formulas

The *wave equation* for small amplitude waves on a thin string of mass per unit length μ and under tension T_0 is

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{u^2} \frac{\partial^2 y}{\partial t^2} \tag{18-1}$$

where u is the velocity of the wave, and is given by

$$u = \left(\frac{T_0}{\mu}\right)^{1/2} \tag{18-2}$$

For a string of length L fixed at its endpoints at $x = 0$ and $x = L$, the only possible frequencies of vibration are

$$\nu_n = \frac{u}{\lambda_n} = \frac{u}{2L} n \qquad (n = 1, 2, \ldots) \tag{18-19}$$

with λ_n the wavelength of the associated normal mode and u as defined in (18-2).

If an observer is in motion at a velocity $\mathbf{v}_0$ ($v_0 < u$) relative to a stationary source of sound (frequency ν_0), then the frequency ν that he hears is

$$\nu = \nu_0\left(1 - \frac{v_0}{u}\cos\theta\right) \tag{18-22}$$

but if the source is in motion at a velocity $\mathbf{v}_s$ ($v_s < u$) and he is at rest, this frequency is

$$\nu = \frac{\nu_0}{1-(v_s/u)\cos\theta} \tag{18-23}$$

QUESTIONS

1. Define or describe briefly the following: (a) wave pulse; (b) dispersive medium; (c) transverse wave; (d) standing wave; (e) destructive interference; and (f) overtone.
2. What is a longitudinal wave and in what way does it differ from a transverse wave? Give two examples of each of these types of waves.
3. Explain in physical terms why sound waves propagating through a gas must be longitudinal. What general characteristics must a medium have so that it can support transverse waves?
4. What is meant by the superposition principle for wave motion? Explain in physical terms the phenomena of constructive and destructive interference by use of this principle.
5. Consider the two wave pulses in Figure 18-6c, polarized at right angles to each other and traveling along the string. Will they interfere with each other when they cross? Is complete destructive interference possible here? Explain.
6. Compare the transfer of energy from point to point in a medium by waves with the corresponding transfer involving the transport of matter. In what outstanding way are these two methods of energy transfer different? That is, what physical effects are associated with energy transfer by waves that are not associated with other methods of energy transfer?
7. Consider a wave pulse propagating on a string. What happens to the velocity u of such a pulse if we increase:
   (a) The length L of the string (μ fixed)?
   (b) The tension T_0 in the string?
   (c) The mass per unit length μ of the string?
8. Suppose that you hold in your hand a tapered whip, whose mass per unit length decreases towards the tip. If you generate a wave pulse by suddenly moving your hand, will the speed of this pulse increase or decrease as it travels towards the tip of the whip? Explain.
9. Consider again the whip of variable mass per unit length in Question 8. As a wave pulse travels toward regions of decreasing mass per unit length, explain why the shape of this pulse will change as the pulse travels along. In what way will the shape of this pulse change?
10. What happens to the speed of the water waves as the water gets shallower? Describe the change in shape (if any) associated with these waves.

Can you account for the existence of breakers at the seashore by use of this effect?

11. Describe an experiment to determine to what extent air is a nondispersive medium for the propagation of sound waves.
12. During a thunderstorm, suppose you see a flash of lightning 5 seconds before you hear the associated thunder. Using the fact that the speed of light is very much greater than that of sound, about how far away from you did the electrical discharge take place?
13. Describe an experiment that you could use to measure the speed of sound in air. Would your proposed method work if the speed of sound were 3000 m/s instead of about 300 m/s?
14. Consider a standing wave on a string. Describe the motion of the element of the string at a node. What is the motion of the element of the string at an antinode?
15. Suppose that, in tuning his instrument, a violinist tightens one of the strings. Does he thereby affect the fundamental frequency of the standing wave on the string? What if he were to loosen it?
16. Explain why it is that when a piano tuner tightens a string he does not change the wavelength of its fundamental mode of vibration, but he does affect the wavelength and the frequency of the sound wave that is heard when the piano is played.
17. If the fundamental of a certain violin string is 440 Hz, what is the frequency of the first overtone? Of the third harmonic?
18. Explain in what way and why the sound that you hear from a certain source sounds different depending on whether the source of the sound recedes from you or approaches you.
19. What is the physical basis for the fact that the frequency of a sound wave emanating from a source approaching you at a velocity **v** is not always the same if the source is at rest and you approach it at the velocity $-\mathbf{v}$?
20. On observing a flying airplane, you note that the sound of its motors does not come from that part of the sky in which the plane is instantaneously located. Does this imply, of necessity, that the plane is flying faster than the speed of sound? Explain.
21. Describe what you would hear if you were approaching a stationary source of sound of frequency ν_0, but at a speed greater than that of sound.
22. A sound source S at rest emits a tone of frequency ν_0 toward a stationary observer O. If a wind blows from S to O, is the frequency that O hears greater or less than ν_0? Explain.
23. A source of sound S emits sound waves of frequency ν_0 when at rest relative to the air. If this source and an observer O both travel in the same direction and at the same speed (less than that of sound), is the frequency ν that the observer hears under these circumstances higher or lower than ν_0? Explain.

PROBLEMS

Unless a statement is made to the contrary, assume in the following problems that the speed of sound in air is 330 m/s.

1. A steel wire 5 meters long has a mass of 0.03 kg. Calculate the velocity of waves on the wire if it is stretched between two points and kept under a tension of 30 newtons.
2. Under what tension must a rope of mass per unit length 0.6 g/cm be kept so that waves will travel along it at a speed of 50 m/s?
3. A long string of length L is stretched

and kept under a tension numerically equal to its weight. Show that the velocity u of transverse waves along this string is

$$u = \sqrt{gL}$$

4. If the observed speed of water waves at a certain part of a lake is 2 m/s, what is the depth of the lake at that point?
5. When observing water waves on the surface of a certain river, suppose you note that the distance between successive crests is 1.2 meters and that in a time interval of 10 seconds, eight crests pass a given point. Calculate **(a)** the speed of these water waves and **(b)** the depth of the water.
6. Show that the velocity of sound in a dilute gas may be expressed in the form

$$u = v_{\mathrm{th}} \left(\frac{\gamma}{3}\right)^{1/2}$$

where γ is the ratio of specific heats and v_{th} is the thermal velocity of the gas molecules.

7. In a laboratory experiment a measurement of a speed of sound in ^{4}He yielded the value 400 m/s. Using the fact that this gas is monatomic, so the ratio of its specific heats has the value of $\gamma = 5/3$, calculate **(a)** the thermal speed of the gas and **(b)** the temperature of the gas, assuming adiabatic conditions.
8. Calculate the speed of sound at 0°C in a dilute gas consisting of a 50-50 mixture of oxygen and nitrogen. Compare your answer with the speed of sound in ordinary air at the same temperature.
9. Consider a wave pulse which at $t = 0$ has the form

$$y(x, 0) = \frac{a^3}{a^2 + x^2}$$

where a is some fixed parameter.

(a) If the pulse travels along the positive sense of the x-axis at a velocity u, what is the displacement y of this wave at any time t?

(b) What would the displacement of the wave have been if it were traveling in the direction of negative values of x?

(c) Show that your solutions to (a) and (b) satisfy the wave equation in (18-1).

10. Consider again the transverse wave on the string in Problem 9. **(a)** Calculate in terms of the parameters a and u the velocity of that particle of the string which instantaneously has the maximum displacement. **(b)** Calculate the acceleration of that particle of the string which is instantaneously a distance a to the right of the point in (a).
11. Show that the following forms of a sinusoidal wave are equivalent to (18-12):

$$y = A \cos 2\pi[(x/\lambda) - (\nu t)]$$
$$= A \cos \omega(t - x/u)$$

12. Suppose that the equation of a traveling wave on a string has the form

$$y = 0.03 \cos \pi(3x - 100t)$$

where y and x are measured in meters and t is in seconds.

(a) What is the amplitude, the wavelength, and the period of this wave?

(b) What is the velocity of the wave?

(c) Calculate the maximum speed of any particle on the string.

13. Consider a sinusoidal wave that travels at a speed of 60 m/s along a string kept under tension. If the amplitude of the wave is 2 cm and its frequency is 100 Hz, write down the equation for this wave, assuming

that it travels along the positive sense of the x-axis.

14. Suppose that for the situation in Figure 18-10 the wire has a mass per unit length of 0.2 g/cm, and is kept under a tension of 50 newtons. If the bob oscillates with a maximum amplitude of 1.2 cm and a period of 0.5 second, calculate **(a)** the velocity of the wave along the string and **(b)** the wavelength λ of this wave motion.
15. Suppose the equation for a wave on a string is

$$y = 3\cos 2\pi\left(\frac{x}{50} - 20t + \frac{\pi}{3}\right)$$

where y and x are in centimeters and t is in seconds.
   (a) Determine the wavelength and the frequency of this wave.
   (b) Calculate the maximum velocity of the particles of the string.
   (c) Determine the maximum acceleration of any element of the string.
16. Write down a formula for the wave that must be added to that in Problem 15 so that the resultant wave is a standing wave. Making use of the trigonometric identity in (18-15), express the total wave in a form analogous to (18-16).
17. For the wire described in Problem 1, calculate **(a)** the frequency ν_1 of the fundamental tone and **(b)** the order of the highest overtone in the audible range ($\leqslant$20,000 Hz).
18. The normal human ear is sensitive to acoustic waves with frequencies in the range 20 Hz to 20,000 Hz. Give the range of wavelengths associated with this frequency interval for **(a)** air at a temperature of 23°C and **(b)** helium at the same temperature.
19. A piano wire has a length of 0.5 m and a mass of 3 grams and is kept under tension of 1000 newtons.
   (a) What is the velocity of waves on the wire?
   (b) What is the frequency of its fundamental mode?
   (c) Calculate the frequency of its second harmonic.
20. The fundamental frequency of a violin string of length 50 cm and of mass 1.5 grams is 500 Hz.
   (a) What is the wavelength of a standing wave on the string when it is vibrating in its fundamental mode?
   (b) Calculate the tension in the wire.
   (c) Where must this wire be "fingered" to change its fundamental frequency to 600 Hz?
21. By what factor must the tension in a piano wire be increased in order to increase the frequency of its fundamental mode of vibration by 25 percent?
22. Suppose that a violin string of length 40 cm vibrates in its fundamental mode at a frequency of 440 Hz.
   (a) What is the wavelength of the wave generated on the string?
   (b) What is the wavelength of the sound wave generated, assuming it to be in air?
   (c) What is the wavelength of the sound generated if it is in an atmosphere of helium at 300 K?
23. Calculate the frequency of the fundamental note generated by an organ pipe of length 20 cm if:
   (a) It is closed at both ends.
   (b) It is open at both ends.
   (c) It is open only at one end.
24. Repeat Problem 23, but assume this time that the pipe is in a helium atmosphere at 23°C.
25. An organ pipe of length 50 cm is closed at one end and is vibrating in its second harmonic. Assuming that the speed of sound in air is 330 m/s, what is the frequency of the tone emanating from the pipe?
26. Consider the sound waves generated by two tuning forks of nearly equal frequencies ν_1 and ν_2 and as

heard by a listener whose distance from the forks is very large compared to their separation. Assume that the respective air-pressure variations p_1 and p_2 on his eardrums may be expressed by

$$p_1 = p_0 \cos 2\pi\nu_1 t \qquad p_2 = p_0 \cos 2\pi\nu_2 t$$

with p_0 a constant.

(a) Show by use of the superposition principle that the total pressure p may be expressed as

$$p = \left\{2p_0 \cos 2\pi\left(\frac{\nu_1 - \nu_2}{2}\right)t\right\} \times \cos 2\pi\left(\frac{\nu_1 + \nu_2}{2}\right)t$$

(b) Explain in what sense this can be viewed as a wave of frequency $(\nu_1 + \nu_2)/2$ and with a slowly varying amplitude at the frequency $(\nu_1 - \nu_2)/2$.

(c) A *beat* is said to occur whenever this amplitude is maximum. Show that the number of beats per second is $(\nu_1 - \nu_2)$.

27. Making use of the result of Problem 26, calculate the number of beats per second that are heard if:

(a) Two tuning forks of respective frequencies 233 Hz and 237 Hz are vibrating.

(b) Two violin strings are composed of the same wire and are under the same tension but one is 1 percent longer than the other. Assume that each vibrates in its fundamental mode and that for the shorter wire $\nu_1 = 550$ Hz.

28. Suppose that the frequency of a locomotive whistle traveling at 35 m/s is 400 Hz. Calculate the frequency heard if instantaneously:

(a) The train approaches the listener.

(b) The train travels at right angles to the line drawn from the locomotive to the listener.

29. Consider again the train in Problem 28, but suppose this time that the train is at rest. Calculate the frequency heard if the listener:

(a) Approaches the locomotive at 50 m/s.

(b) Recedes from it at 50 m/s.

30. With what minimum speed must a source that vibrates at 12,000 Hz approach an observer so that he hears nothing? Assume that 20,000 Hz is the maximum audible frequency.

31. Show that for low velocities, $v_0 = v_s \ll u$, (18-22) and (18-23) correspond to the same frequency shift. Does this mean that in this limit the Doppler shift depends only on the relative motion between the source and the observer?

32. At what angle with respect to the horizontal will a jet be when you first hear its engine if it is traveling at:

(a) 660 m/s?

(b) Mach 3?

33. At what speed is a jet traveling if you first hear the sound from its engines when its position is at an angle of 20° with respect to the vertical?

34. If the angle between the wavefronts in the wake of the boat in Figure 18-18 is 60°, and the boat is traveling relative to water at a speed of 10 m/s, calculate the depth of the water there.

†35. For the physical situation in Problem 12, calculate the maximum potential energy per unit length of an element of the string. Assume $\mu = 0.1$ kg/m.

†36. For the wave in Problem 15, calculate the maximum value for the kinetic energy and the potential energy per unit length of the string. Assume $\mu = 0.2$ kg/m.

†37. Prove that the energy E in (18-25) is constant in time if the endpoints of the string are fixed.

19 Coulomb's law

From a long view of the history of mankind—seen from say ten thousand years from now—there can be little doubt that the most significant event of the 19th century will be judged as Maxwell's discovery of the laws of electrodynamics.

R. P. FEYNMAN

19-1 General introduction

For the next eleven chapters we shall be studying the branch of physics known as *electromagnetism*. As the name implies, this discipline deals with electric and magnetic phenomena and the relations between them. Although some of the qualitative observations on electricity and magnetism date back to antiquity, most of the key experimental and theoretical ideas were discovered by researchers who lived during the nineteenth century. Very prominent among these is James C. Maxwell (1831–1879), who succeeded in a synthesis of all experimental facts on electromagnetism known up to about 1860—and since discovered—in terms of a set of four equations, which today are known as *Maxwell's equations*. These very fundamental relations play a role in electromagnetism which is very analogous to that played by Newton's laws in mechanics. By use of these equations we are able to understand not only many aspects of the forces between atoms and molecules but also the principles underlying the operation of a variety of

devices, including electric motors, radio and television transmitters and receivers, and high-energy particle accelerators such as cyclotrons and synchrotrons. The scope of these laws is obviously enormous.

For the earliest recorded observations of electric and magnetic phenomena we are indebted to the Greeks. The fact that rubbed amber acquires electrical properties, as evidenced by the fact that it attracts small pieces of straw, was recorded by Thales of Miletus circa 600 B.C. Similarly, the existence of lodestone (that is, "leading stone" or compass) appeared in Greek writings as early as 800 B.C., and the magnetic properties of magnetite (the iron ore consisting mainly of $FeO\text{-}Fe_2O_3$) was known to Pliny. From these very early and primitive observations, it took man more than 2000 years to establish the fact that electricity and magnetism are but different aspects of the same phenomenon. For this breakthrough it was necessary to await the development of the voltaic cell, or battery, which made possible the production of steady electric currents. In 1820, Hans Oersted (1777–1851) discovered that a wire through which such an electric current flows has properties similar to that of a magnet. Shortly thereafter, Michael Faraday (1791–1867) reported on his observations of the related effect that if a wire is moved near a magnet—or, equivalently, if a magnet is moved near a loop of wire—an electric current flows in the wire. Finally, Maxwell showed that these and a variety of other experimental facts could be correlated in terms of a small number of simple relations.

One of the very interesting by-products of Maxwell's formulation was his deduction that if the current flow in a wire varies in time, then waves would be radiated. Twenty years after Maxwell's enunciation of his theory, the existence of these *electromagnetic waves* was established experimentally by Heinrich Hertz (1857–1894). From these early observations of Hertz it was a relatively small step to develop "wireless" transmission of electromagnetic signals and to establish the fact that ordinary light is but one example of an electromagnetic wave. Indeed, as we shall see, it is possible to deduce the speed of light by making measurements of electromagnetic phenomena in the laboratory!

As a final note, toward the end of the nineteenth century it became apparent that there were certain logical inconsistencies between Maxwell's equations and Newton's laws of motion. This matter was unambiguously resolved by Albert Einstein (1879–1955) in 1905 when he enunciated his theory of relativity. Remarkably enough, Maxwell's equations withstood the test. Newton's laws did not and required modification.

19-2 Electric charge

One of the simplest ways to produce electric charge involves a hard-rubber object, such as a pocket comb. If you comb your hair vigorously with such a comb you will find that it is able to attract small objects such as pieces of

paper. This is analogous to the ancient Greek observation that rubbed amber attracts bits of straw. More generally, experiment shows that two hard-rubber rods which have been rubbed with animal fur will repel each other. Similarly, if two glass rods are rubbed with silk, they will also be found to repel each other. However, if a hard-rubber rod that has been rubbed with fur is brought near a glass rod that has been rubbed with silk we find that they attract each other. In all three of these cases, the force between the rods decreases as the separation distance is increased.

A convenient way to demonstrate the existence of these forces is shown in Figure 19-1. A glass rod that has been rubbed with silk is suspended by a nonmetallic thread in such a way that it is free to rotate in a horizontal plane. If a hard-rubber rod which has been rubbed with fur is brought near the glass rod, it will be observed to rotate, thus demonstrating the existence of an attractive force between the rubbed parts of the rods. If the suspended rod in the figure is replaced by a rubber rod, the force is found to be repulsive.

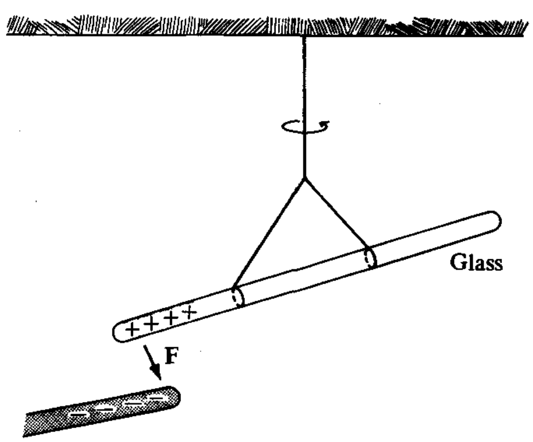


Figure 19-1

More generally, experiment shows that when rubbed in an appropriate way, many solid substances will behave either as the glass rod or as the hard-rubber rod. That is, they will be either attracted by the glass rod and repelled by the hard-rubber rod, or vice versa. We characterize this property of attraction or of repulsion of a rubbed solid relative to a glass or a rubber rod by saying that these bodies have an *electric charge.* It is apparent from the above discussion that there exist at least two types of electric charge. Following the convention originally introduced by Benjamin Franklin, we call the type of electric charge on a glass rod rubbed with silk as *positive* and that on a rubber rod rubbed with fur as *negative.* In operational terms, then, any substance that repels a glass rod (and thus attracts a hard-rubber rod) is said to have a positive charge, and conversely, any substance that repels a hard-rubber rod (and thus attracts a rubbed glass rod) is said to have a negative charge. The plus and minus signs on the rods in Figure 19-1 reflect this universally accepted convention.

These results on the electric forces between charged solids may be conveniently summarized by the statement:

Positively charged bodies repel other positively charged bodies and negatively charged bodies repel other negatively charged bodies, whereas a body carrying either sign of electric charge will attract a body carrying charge of the opposite sign.

Let us now reexamine the above experimental results involving macroscopic bodies in light of our current understanding of atoms and molecules. Studies in chemistry and atomic physics have established that an atom consists of a very massive, positively charged nucleus, about which orbit a number of relatively light, negatively charged electrons. Under ordinary circumstances the atom is electrically neutral in that it carries no net charge. If, however, a negatively charged electron is torn from its parent nucleus, then two electric charges come into existence: a positively charged ion and the negatively charged electron itself. In terms of electrons and ions then, when a glass rod is rubbed with silk, some electrons or some negatively charged ions must come off during the rubbing process. In this way it acquires a positive charge. Correspondingly, the fact that a hard-rubber rod when rubbed with animal fur acquires a negative charge must mean in microscopic terms that electrons are added to the rod. Although it would be difficult, at this point, to justify this microscopic picture, it is fully correct and is a convenient way for thinking of all electric and magnetic phenomena. It will be freely used in the following discussions.

Unfortunately, the above experiments involving rubbed glass and rubber rods do not lend themselves readily to quantitative and reproducible experiments. The precise strength of the force between various electrically charged, macroscopic bodies depends sensitively on the detailed geometry of the bodies, on the length of time they are rubbed, and so forth. Thus to go from these qualitative observations to a quantitative measure of the forces between charged bodies it is convenient to make use of a class of materials known as conductors.

19-3 Charge induction and conductors

In a discussion of electric phenomena, it is convenient to characterize matter as being either a *conductor* or an *insulator*.[1] A conductor is a substance through which electric charge is readily transported, whereas an insulator is

[1]A third class of substances, known as *semiconductors*, lies somewhere between these two. In the following we shall be concerned only with conductors and insulators.

one which will not easily conduct charge. For example, the hard-rubber rods and the glass rods considered in Section 19-2 are insulators, for experiment shows that only the rubbed portions of the rods acquire an electric charge; there is no tendency for this charge to leave the surface and to diffuse into the interior of the rods. By contrast, if charge is placed on a metallic conductor, such as copper or silver, the electric charge appears not only on that portion of the conducting surface on which it is placed, but at various other parts of the conductor as well.

In addition to the familiar metallic solids, such as copper or silver, conductors also occur naturally in the form of liquids and gases. A *plasma* is a gas that consists of mobile charged particles and is thus a gaseous conductor. Plasmas can be produced in the laboratory; they also occur naturally in the upper regions of our atmosphere (the ionosphere), in the region surrounding the sun, and generally throughout much of intergalactic space. As exemplified by mercury, which at room temperature is a liquid, conductors can also exist in the liquid state. In addition to liquid metals, water containing various salts in solution, such as AgCl and Na_2CO_3, are also very good conductors. They are known as *electrolytes*. The human body itself is a relatively good conductor by virtue of the body fluids. The problem of describing the electrical behavior of gaseous and liquid conductors is generally much more complex than is the corresponding description of solid ones, and thus from now on we shall deal exclusively with solid conductors.

In connection with a discussion of electric charge in solid conductors, it is helpful to have available a microscopic picture of these materials. For this purpose we may think of a conductor as consisting of a rigid array of immobile and positively charged ions, which are arranged in a regular lattice of some type. Figure 19-2 shows a two-dimensional version of such a lattice. Interspersed between the fixed ions is an isotropically distributed gas of negatively charged electrons, with typically one or two electrons for each

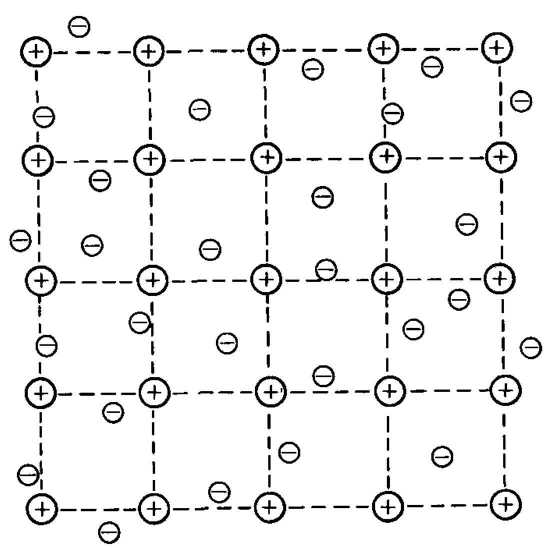


Figure 19-2

ion. In the normal situation of electrical neutrality, these electrons are distributed uniformly, so that in any small, but macroscopic, volume element of the lattice there is just as much positive charge associated with the ions as there is negative charge on the electrons. The lattice is in this case said to be *electrically neutral.* Note the important point that the positively charged ions are "rigidly" held in place but that the electrons are very mobile. This means that if electric charge is brought near a conductor, even though both the electrons and the ions will experience an additional force, only the electrons are free to move under its action. For simplicity we shall always think of the ions as maintaining fixed relative positions, although in actual fact they also undergo thermal motions about their equilibrium positions.

Figure 19-3a shows a charged glass rod near a conductor. According to the above physical picture, some of the electrons will be attracted by the electric charge on the rod and will move toward that portion of the conducting surface closest to the glass rod. As a result of this migration, a positive charge will appear on the more distant surface of the conductor. (It will be established in the next chapter that charge can only reside on the outer surface of a solid conductor.) Similarly, if a rubbed hard-rubber rod with its negative charge is brought near a conductor, some of the mobile electrons will be repelled by the rod, and the situation will then be as shown in Figure 19-3b. Hence, if charge is brought near a conductor an *induced charge* will appear on the surface of that conductor. Further, since the force between charged bodies decreases as the distance between them increases, the force between the conducting sphere and the rod, in Figure 19-3, is attractive. In both cases, the sign of the induced charge on that portion of the conducting surface nearest to the rod is opposite to the sign of the charge on the rod; thus there is a net attractive force between them.

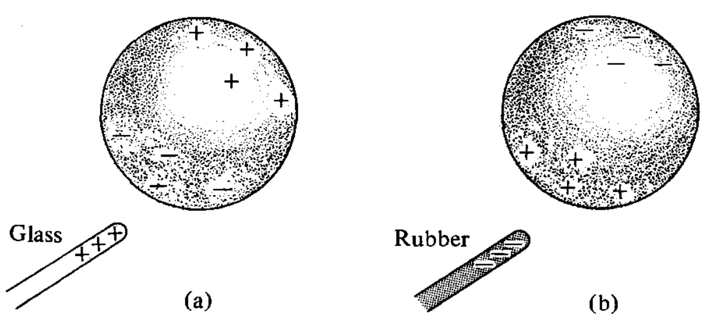


Figure 19-3

19-4 Grounds

A concept of importance in a study of the electrical behavior of conductors is that of a *ground.* To introduce this notion imagine bringing up a charged hard-rubber rod to a conductor A, which is very far away from and connected to a second one B by a long conducting wire (Figure 19-4). As

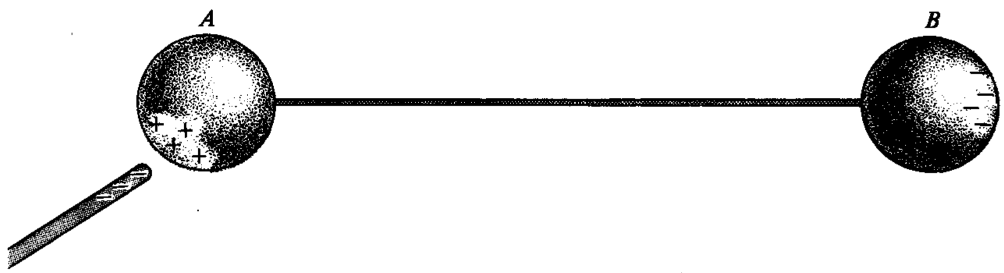


Figure 19-4

before, some of the electrons on A will respond to the repulsive force exerted by the negatively charged rod and will attempt to travel as far away from this charge as possible. Unlike the situation in Figure 19-3, however, where the electrons could not cross the boundary surface of the isolated conductor, in the present case, because of the connecting wire, these electrons can very easily leave A and travel to conductor B. We say that conductor A has been *grounded* provided that the second conductor B is so far away that the charge on it no longer influences the situation prevailing in the neighborhood of A. The symbol

⏚

will be used to denote the fact that a conductor has been grounded. If the conducting wire in Figure 19-4 is sufficiently long so that B is very far away from A and is thus grounded, then we represent this situation as in Figure 19-5.

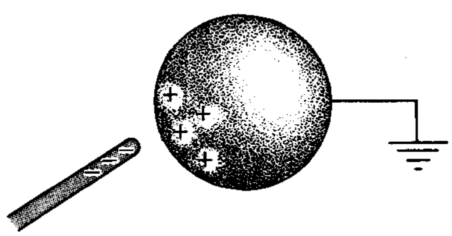


Figure 19-5

Mainly because of its metallic content and the existence of dissolved salts in the oceans, the earth itself is a relatively good conductor. Furthermore, because of its immense size it is in effect an infinite source as well as a sink of charge and is thus also a ground! Indeed, the name *ground* literally means the connecting of a conductor, by means of a conducting wire, to the earth. If we think of conductor B in Figure 19-4 as being the earth itself, then the negative charge appearing on it is always far away from conductor A and thus, even if the conducting wire connecting them is not particularly long, A will invariably be grounded under these circumstances.

As an application of the notion of grounding let us see how it can be used to place charge on a conductor. This process is known as *inducing charge* on a conductor or *charging it by induction.* Suppose, in Figure 19-6a, that a negatively charged rubber rod is brought up to an isolated conductor so that,

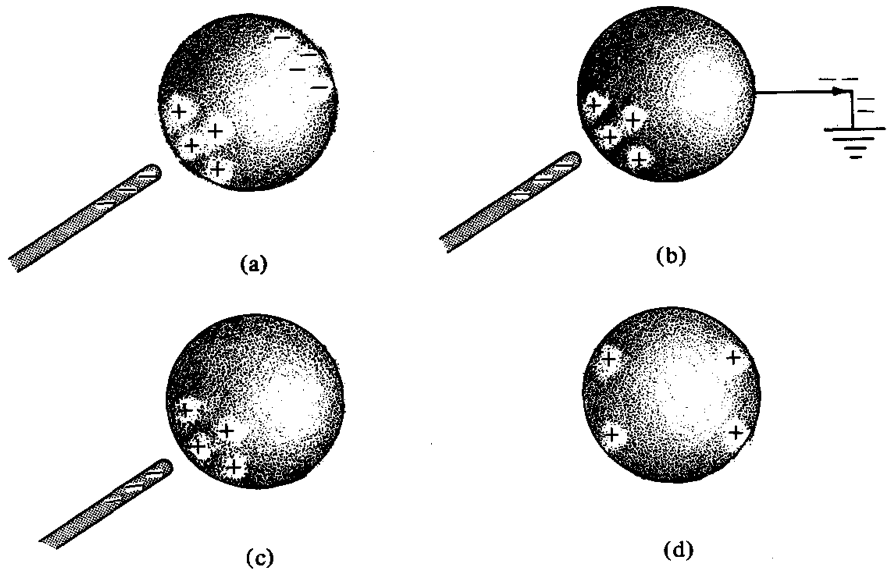


Figure 19-6

as shown, some of the electrons migrate to that side of the conductor farthest from the rod. If this conductor is now grounded, then some of the electrons will leave the original conductor and will no longer be of any direct interest. If the connection with the ground is then cut, the conductor is again isolated. Finally, as shown in Figures 19-6c and 19-6d, by removing the hard rubber rod, we obtain a conductor with a net positive charge on it.

Similarly, by use of a rubbed glass rod, a conductor containing a net negative charge can be produced.

19-5 Coulomb's law—qualitative aspects

To obtain a quantitative measure of the force between charged bodies, it is convenient to make use of small, conducting spheres. We shall say in this connection that two spheres are *small*, provided their radii are negligible compared to the distance separating them. In effect, then, small, conducting spheres may be thought of as charged particles, and in the following the term "particle" will always be used in this sense.

We shall now describe certain experiments that establish the following properties of the force between two charged particles:

1. The direction of the force lies along the line joining them.
2. The magnitude of the force varies inversely as the square of the distance between them.
3. The force is directly proportional to the charge on each particle.
4. For a given separation distance, the magnitude of the force, although not its sense, is independent of the sign of the charges.

Consider, in Figure 19-7, two charged particles, A and B, which have the respective masses m_1 and m_2 and are suspended by two insulating strings of

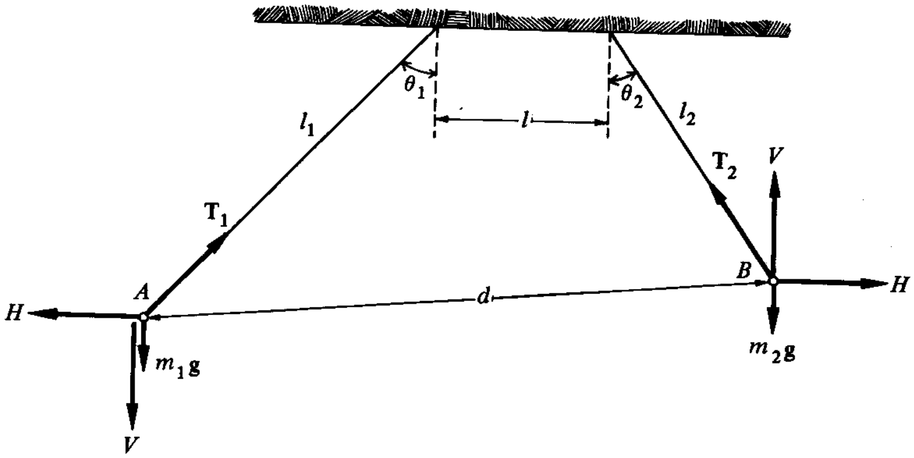


Figure 19-7

lengths l_1 and l_2 from a horizontal ceiling. Assuming, to be specific, that the particles have the same sign of charge, there is a repulsive force between them, and at equilibrium the strings will hang at certain angles, θ_1 and θ_2, with respect to the vertical. Besides their weights, m_1g and m_2g, each particle also experiences a force due to the tension in the strings (T_1 and T_2 in the figure) as well as a force **F** due to the electric repulsion between them. For convenience, only the horizontal and vertical components, H and V, respectively, of this electric force **F** have been drawn in the figure. Also, we have made use of Newton's third law, according to which the electric force on one of the particles is equal and opposite to that on the other.

Since both particles A and B are in static equilibrium under the combined action of these forces, it follows from the principles of mechanics that

$$\begin{aligned} -V + T_1 \cos\theta_1 - m_1 g &= 0 \\ T_1 \sin\theta_1 - H &= 0 \\ T_2 \cos\theta_2 + V - m_2 g &= 0 \\ H - T_2 \sin\theta_2 &= 0 \end{aligned} \qquad (19\text{-}1)$$

where the first two equations are the respective sums of the vertical and the horizontal components of the force acting on A, and the second two represent the corresponding quantities for B. If we eliminate the unknown tensions T_1 and T_2, the resultant two equations for the components V and H of the electric force may be solved in terms of the masses m_1 and m_2 and the two angles θ_1 and θ_2. Hence, since the four parameters m_1, m_2, θ_1, and θ_2 are easily measured independently, it follows that the strength of the electric force between the two particles can be obtained directly.

In terms of the setup in Figure 19-7, it is a straightforward matter to confirm the above properties of the electric force between two charged particles. By varying the distance l separating the points of suspension of the supporting strings, we can in effect vary the separation distance d between the particles. Carrying out a sequence of such experiments we find

in this way that the magnitude F of the electric force $\mathbf{F}$ varies inversely as the square of the separation distance d; that is,

$$F = (V^2 + H^2)^{1/2} \propto \frac{1}{d^2}$$

Further, since V and H are also separately measurable, the direction of $\mathbf{F}$ can also be measured in this way. The result is that $\mathbf{F}$ invariably lies along the line joining the two particles. Thus the first two properties of the electric force between two charged particles, as listed above, are established.

To confirm the fact that $\mathbf{F}$ is also proportional to the charge on each particle, we proceed as follows. Assuming, to be specific, that the charges on A and B in Figure 19-7 are positive, let us take, say, B and as in Figure 19-8 place it into electrical contact with a second originally uncharged and *identical* conducting sphere B'. It follows from symmetry that these two identical spheres must share the available electric charge equally; thus after the spheres are separated, the charge on B will be precisely half of its original value. If the experiment in Figure 19-7, with A still having its original charge, is now repeated, but with the charge on B decreased by a factor of two, we find that for any fixed value for d, the electric force $\mathbf{F}$ between them is halved. In this way, then, by varying the amount of charge on both spheres we may confirm that, in general, the force between two charged particles varies directly as the product of their charges. It is interesting to note that it is possible by use of this symmetry argument to deduce this direct proportionality without having specified a unit of electric charge. The question of units will be considered in Section 19-6.

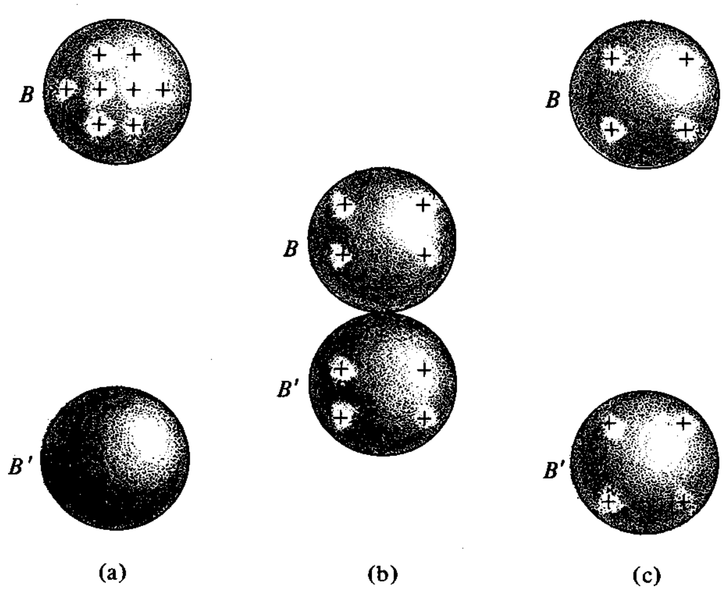


Figure 19-8

Finally, to confirm the fact that the magnitude of the force is independent of the sign of the charge, let us place equal, but opposite, charges on two

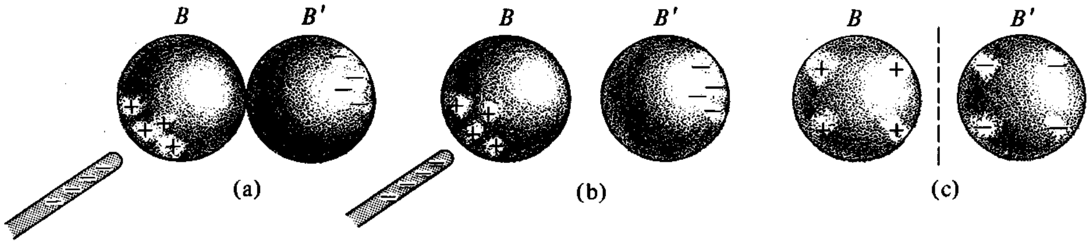


Figure 19-9

identical conducting spheres, B and B'. This may be done by placing the two conducting spheres into electrical contact, and then bringing up a charged rubber rod so that, as in Figure 19-9a, electrons migrate from B to B'. If the two spheres are separated and the rod is removed, then as shown in Figures 19-9b and c, the two conducting spheres will have equal and opposite charges. With a fixed charge on sphere A in Figure 19-7 we now carry out two successive experiments: one with sphere B, with its positive charge, and one with sphere B', with its equal but opposite negative charge. In this way, the fact that the magnitude of the force is independent of the sign of the charge is readily confirmed.

It should be noted that there are difficulties of a practical nature associated with the usage of the apparatus in Figure 19-7, and the above experiments should be thought of only as "thought experiments." Indeed, in ascertaining the properties of the force between charged particles, Coulomb (1736–1806) did not use this apparatus. Instead he used an apparatus similar to that shown in Figure 4-13, which was used by Cavendish to measure the gravitational constant G and is known as a torsion balance. Nevertheless, the experiments above are easy to understand and in principle may be carried out and used to justify quantitatively the stated properties of the electric force between small, charged bodies.

19-6 Coulomb's law—quantitative aspects

Based on the experiments of Section 19-5, we know that the electric force between two charged particles varies directly with the product of the charges and inversely with the square of the distance between them. In quantitative terms this means that if two particles of charges q and Q (to be defined below) are separated by a distance r, then the force $\mathbf{F}$ between them has the magnitude

$$F = k\frac{qQ}{r^2} \tag{19-2}$$

and its direction lies along the line joining the particles. As shown in Figure 19-10, if the charges have the same sign—that is, if they are either both positive or both negative, so that $Qq > 0$—then the force is repulsive. Correspondingly, if they are of the opposite sign, so that $Qq < 0$, then the

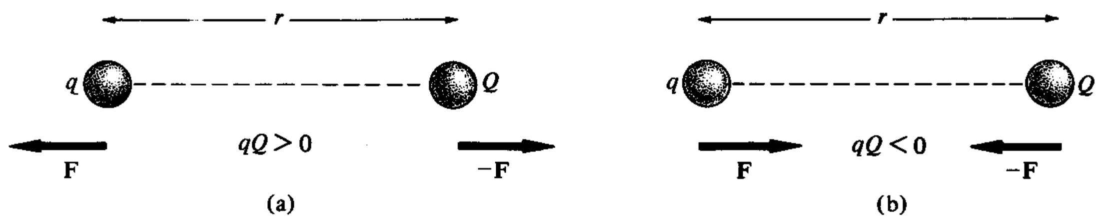


Figure 19-10

force is attractive. The proportionality constant k in (19-2) is arbitrary, and its detailed specification depends on the system of units adopted.

Although it is possible to do so, for technical reasons having to do mainly with the accuracy of experiments, it is *not* convenient to define a unit of charge in terms of (19-2) directly. As will be discussed in Chapter 26, it is more convenient first to define the SI unit of electric current of the *ampere* and then to define the associated unit of electric charge of the *coulomb* (C) as the amount of charge that is transported in 1 second along a wire through which flows a current of 1 ampere. A precise definition of the coulomb must therefore await our discussion of electric currents. In the interim, however, we shall make free use of this unit of electric charge as well as the related unit of the microcoulomb (μC), which is defined to be 10^{-6} coulomb. For purposes of orientation, let us note that the magnitude of the charge on a proton or an electron is 1.6×10^{-19} coulomb $= 1.6 \times 10^{-13}$ μC.

Once the unit of electric charge is established, the proportionality constant k in (19-2) is uniquely determined by experiment. It is customary in the SI system of units to introduce the symbol $1/(4\pi\epsilon_0)$ for this constant. Thus Coulomb's law in (19-2) becomes

$$F = \frac{1}{4\pi\epsilon_0}\frac{qQ}{r^2} \tag{19-3}$$

The constant ϵ_0 is called the *permittivity of free space*, and the factor 4π in this formula is for future convenience. In SI units, where F is measured in newtons, r is measured in meters, and q and Q are in coulombs, the constant $1/(4\pi\epsilon_0)$ has the experimental value[2]

$$\frac{1}{4\pi\epsilon_0} = 8.9875 \times 10^9 \frac{\text{N-m}^2}{\text{C}^2} \tag{19-4}$$

and the associated value for ϵ_0 itself is

$$\epsilon_0 = 8.8542 \times 10^{-12} \frac{\text{C}^2}{\text{N-m}^2}$$

For our purposes the approximate values

$$\frac{1}{4\pi\epsilon_0} = 9.0 \times 10^9 \frac{\text{N-m}^2}{\text{C}^2} \qquad \epsilon_0 = 8.9 \times 10^{-12} \frac{\text{C}^2}{\text{N-m}^2}$$

are usually adequate.

[2]More accurately, $10^7/4\pi\epsilon_0 \equiv c^2$, where c is the speed of light in vacuum.

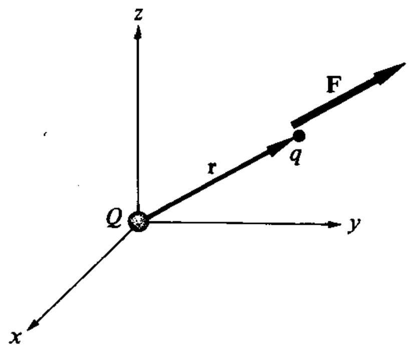


Figure 19-11

Even though force is a vector, in (19-3) the direction of action of this force is not explicitly stated. When making use of (19-3) it is always understood that this force is either attractive or repulsive, depending on the relative signs of the charges, and acts along the line joining the two particles.

For some purposes, it is convenient to rewrite Coulomb's law in its vector form. Suppose, in Figure 19-11, that the origin of a coordinate system is selected at the position of the particle of charge Q and that **r** represents the vector which describes the location of the other particle. In terms of these quantities, Coulomb's law for the force **F** acting on the particle of charge q is

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} qQ \frac{1}{r^3} \mathbf{r} \tag{19-5}$$

since $\mathbf{r}/r$ is a unit vector parallel to the direction of **r**. If the particles have charges of the same sign, then they repel each other, and the force **F** in this formula points along the direction of **r** as in the figure. On the other hand, if the charges are of opposite sign, then, according to (19-5), **F** points in the direction $-\mathbf{r}$. This is consistent with the fact that the particles attract each other in this case.

Example 19-1 Calculate the magnitude of the force of attraction between an electron and a proton separated by a distance of 0.5×10^{-10} meter, as they are in a hydrogen atom. Compare this with the force of attraction due to the action of gravity.

Solution We are given the data

$$q = -1.6 \times 10^{-19}\,\mathrm{C} \qquad Q = +1.6 \times 10^{-19}\,\mathrm{C} \qquad r = 0.5 \times 10^{-10}\,\mathrm{m}$$

Since the charges are of opposite sign, the force is attractive, and its magnitude F according to (19-3) is

$$F = \frac{|qQ|}{4\pi\epsilon_0} \frac{1}{r^2} = 9.0 \times 10^9 \frac{\mathrm{N\text{-}m^2}}{\mathrm{C^2}} \times \frac{(1.6 \times 10^{-19}\,\mathrm{C})^2}{(0.5 \times 10^{-10}\,\mathrm{m})^2}$$
$$= 9.2 \times 10^{-8}\,\mathrm{N}$$

The gravitational force F_G between the electron and proton is given, according to Newton's law of universal gravitation, by

$$F_G = \frac{GMm}{r^2}$$

where $G = 6.67 \times 10^{-11}$ N-m^2/kg^2 is the gravitational constant, $m = 9.1 \times 10^{-31}$ kg is the electron mass, and $M = 1.67 \times 10^{-27}$ kg is the proton's mass. Substituting these values and using $r = 0.5 \times 10^{-10}$ meter, we obtain for F_G the value

$$\begin{aligned} F_G &= 6.67 \times 10^{-11} \frac{\text{N-m}^2}{\text{kg}^2} \times \frac{(9.1 \times 10^{-31}\ \text{kg}) \times (1.67 \times 10^{-27}\ \text{kg})}{(0.5 \times 10^{-10}\ \text{m})^2} \\ &= 4.1 \times 10^{-47}\ \text{N} \end{aligned}$$

so the electric force between these particles is larger than the gravitational force by a factor of about 10^{40}! It is for this reason that the gravitational force between elementary particles can usually be neglected.

Example 19-2 Suppose the charge on the electron were not precisely the same in magnitude as that on the proton but, say, had the value $-(1 - 10^{-6}) = -0.999999$ times that of the proton.

(a) What would be the net charge associated with 1 mole of monatomic hydrogen?

(b) What would be the force of repulsion between 2 moles of hydrogen at a separation distance of 1 meter under these circumstances?

Solution

(a) In terms of Avogadro's number $N_0 = 6.0 \times 10^{23}$ atoms/mole, the charge on the protons is

$$N_0 \times 1.6 \times 10^{-19}\ \text{C}$$

whereas that on the electrons would be

$$-N_0 \times (1 - 10^{-6}) \times 1.6 \times 10^{-19}\ \text{C}$$

Thus the net charge Q_0 on 1 mole of hydrogen under these circumstances is

$$\begin{aligned} Q_0 &= N_0 \times 1.6 \times 10^{-19} (1 - 1 + 10^{-6})\ \text{C} \\ &= 6.0 \times 10^{23} \times 1.6 \times 10^{-19} \times 10^{-6}\ \text{C} \\ &= 9.6 \times 10^{-2}\ \text{C} \end{aligned}$$

(b) According to Coulomb's law, the force F of repulsion between two charges $Q_0 = 9.6 \times 10^{-2}$ coulomb at a separation distance of 1.0 meter is

$$\begin{aligned} F &= \frac{Q_0^2}{4\pi\epsilon_0} \frac{1}{r^2} = 9.0 \times 10^9 \frac{\text{N-m}^2}{\text{C}^2} \times \frac{(9.6 \times 10^{-2}\ \text{C})^2}{(1.0\ \text{m}^2)} \\ &= 8.3 \times 10^7\ \text{N} \end{aligned}$$

This is an enormous force and corresponds to approximately 10^4 tons!

19-7 Coulomb's law for collections of charged particles

Experiments analogous to those described in Figure 19-7, but involving a collection of more than two charged particles, show that the electric force acting on any one particle is the vector sum of the forces that would act on it if each of the other particles were the only other one present. For the case of three particles, for example, this means that the force on particle 1 is the vector sum of the forces produced on it if only particles 1 and 2 were present plus the force produced on it if only 1 and 3 were present. Thus to calculate the electric force on a charged particle in the presence of two or more others, we simply calculate the force due to each one by use of (19-5) and then add the results together vectorially.

As a special case consider the situation in Figure 19-12 of three particles of charges q_1, q_2, and q_3 of the same sign and separated by the distances r_{12}, r_{13}, and r_{23}, respectively. The force $\mathbf{F}_{12}$ on q_1 due to the presence of q_2 is directed as shown in the figure and has the magnitude

$$F_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}^2}$$

Correspondingly, the force $\mathbf{F}_{13}$ on q_1 due to q_3 lies along the line joining q_1 and q_3 and has the magnitude

$$F_{13} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_3}{r_{13}^2}$$

The total force $\mathbf{F}_1$ on q_1 is, as shown in the figure, the vector sum of these:

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} \qquad \textbf{(19-6)}$$

The generalization of this formula to more than three charged particles is obvious and is illustrated in the examples which follow.

The relation in (19-6), which states that the force on a given particle is the vector sum of those produced by the other particles, is known as the *superposition principle.* Strictly speaking, this principle is valid only if the

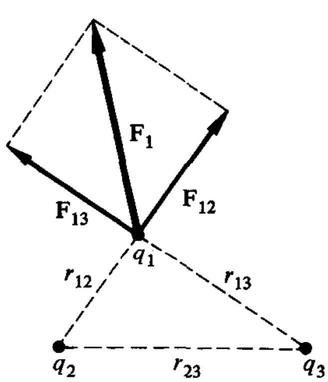


Figure 19-12

"particles" are actually point particles. For if they were, say, charged conducting spheres, then, as will be illustrated below, the electric force acting between spheres 1 and 2 will, in general, be different depending on the precise location of the third sphere. This follows from the fact that the presence of sphere number 3 will, in general, cause a redistribution of the electrons in spheres 1 and 2 and thus modify the electric force that the latter two exert on each other. Since in writing down (19-6) we have not made allowance for this possibility, this formula is of necessity restricted to point charges.

To illustrate this limitation on Coulomb's law, we shall now show that the force between two conducting spheres, only one of which carries an electric charge, is not zero as it would be for point particles. Consider, in Figure 19-13a, two conducting spheres A and B, which initially are separated by a distance much greater than their radii, and suppose that A has zero electric charge and B has a positive charge $Q \neq 0$. Since the spheres are separated by a very large distance, we may treat them effectively as point particles. Thus it follows from Coulomb's law that there is no force between them.

Let us now bring them closer together. As illustrated in Figure 19-13b, because of the positive charge on B some of the electrons in A will migrate to the side of the sphere closest to B. In other words, the presence of charged sphere B causes a redistribution of the charge in A. Because of the resultant charge separation and the inverse-square nature of Coulomb's law, it follows that there will now be a certain force of attraction **F** between the two spheres. Thus, even though A is overall electrically neutral, since B has a net charge and the separation distance between the spheres is comparable to their radii, they will attract each other. Of course, there is no real contradiction with (19-6), since it is valid only for point particles.

A similar failure of the superposition principle for large bodies can be seen in the following setup. Consider first the two uncharged conducting spheres, A and B, in Figure 19-14a. Since neither carries a charge, no charge redistribution takes place in either one, and there is no force between them. If, now, as in Figure 19-14b, a particle of, say, positive charge q is brought near, a rearrangment of charge within the two spheres takes place and they repel each other. That is, not only will each sphere experience a force of attraction due to the charged particle, but in addition each sphere will experience a force of repulsion due to the charge separation in the other

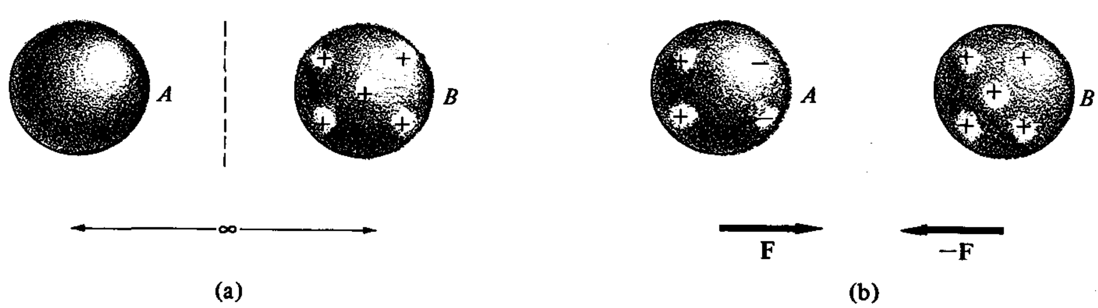


Figure 19-13

Figure 19-14

sphere. In other words, the presence of the charged particle causes an alteration in the force between the conducting spheres, again in seeming violation of the superposition principle. Note that this failure of superposition is due to the usage of bodies of finite dimensions; no such difficulties arise for point particles. In the following discussions we shall continue to restrict ourselves to a consideration of charged point particles; for these, Coulomb's law and the principle of superposition are always valid.

Example 19-3 Two negatively charged electrons, each of charge $-q$, are arranged as in Figure 19-15 near a positively charged alpha particle of charge $2q$. Assuming that $q = 1.6 \times 10^{-19}$ coulomb and $a = 1.0 \times 10^{-10}$ meter, calculate:

(a) The force on one of the electrons.

(b) The force on the positively charged alpha particle.

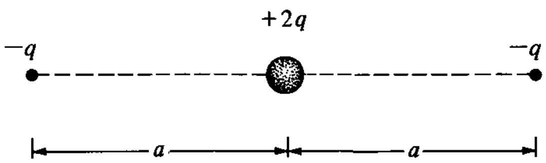


Figure 19-15

Solution Since the three particles lie along a straight line, it follows that all electric forces must also act along this direction. The force of repulsion F_r between the two electrons is given, according to Coulomb's law, by

$$F_r = \frac{q^2}{4\pi\epsilon_0}\left(\frac{1}{2a}\right)^2 = 9.0 \times 10^9 \frac{\text{N-m}^2}{\text{C}^2} \times \frac{(1.6 \times 10^{-19}\,\text{C})^2}{(2 \times 10^{-10})^2}$$
$$= 5.8 \times 10^{-9}\,\text{N}$$

The attractive force F_a between the alpha particle and either electron is

$$F_a = \frac{(+q)(2q)}{4\pi\epsilon_0}\frac{1}{a^2} = 9.0 \times 10^9 \frac{\text{N-m}^2}{\text{C}^2} \times \frac{2 \times (1.6 \times 10^{-19}\,\text{C})^2}{(1.0 \times 10^{-10}\,\text{m})^2}$$
$$= 4.6 \times 10^{-8}\,\text{N}$$

(a) The force on the electron on the right in Figure 19-15 is equal to the repulsive force $\mathbf{F}_r$ acting to the right plus the attractive force $\mathbf{F}_a$ due to the α particle acting to the left. Thus, the net force acting on it is to the *left* and has the magnitude

$$F = F_a - F_r = 4.6 \times 10^{-8}\,\text{N} - 5.8 \times 10^{-9}\,\text{N} = 4.0 \times 10^{-8}\,\text{N}$$

The force acting on the other electron has the same magnitude but acts to the right.

(b) The force on the α particle vanishes! This follows since the attractive force on it due to the electron on the right is just compensated for by the attractive force due to the other electron.

Example 19-4 Four particles, each of charge q, are located at the vertices of a square of side a. Calculate the force on one of these particles.

Solution The situation is shown in Figure 19-16. To calculate the force on, say, the particle located at the upper right-hand vertex, let us set up a Cartesian coordinate system as shown. Then the force $\mathbf{F}_1$ due to the particle at the lower right-hand vertex is

$$\mathbf{F}_1 = \frac{q^2}{4\pi\epsilon_0}\frac{1}{a^2}\mathbf{j}$$

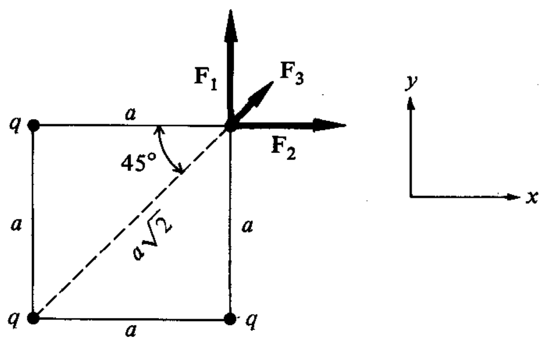


Figure 19-16

where $\mathbf{j}$ is a unit vector along the y-axis. Similarly, the force $\mathbf{F}_2$ due to the particle at the upper left-hand vertex is

$$\mathbf{F}_2 = \frac{q^2}{4\pi\epsilon_0}\frac{1}{a^2}\mathbf{i}$$

Finally, the force $\mathbf{F}_3$ due to the particle at the third vertex, which is at a distance $\sqrt{2}a$ from the given particle, is

$$\mathbf{F}_3 = \frac{q^2}{4\pi\epsilon_0}\frac{1}{(\sqrt{2}a)^2}(\mathbf{i}\cos 45° + \mathbf{j}\sin 45°)$$

since the vector $(\mathbf{i}\cos 45° + \mathbf{j}\sin 45°)$ is a unit vector making equal angles with the x- and y-axes.

Collecting these results together, we find that the total force $\mathbf{F}$ is

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = \frac{q^2}{4\pi\epsilon_0}\frac{1}{a^2}\left[\mathbf{i}\left(1+\frac{\sqrt{2}}{4}\right)+\mathbf{j}\left(1+\frac{\sqrt{2}}{4}\right)\right]$$

since

$$\cos 45° = \sin 45° = \frac{1}{2}\sqrt{2}$$

The magnitude of $\mathbf{F}$ is

$$F = \left(\sqrt{2}+\frac{1}{2}\right)\frac{q^2}{4\pi\epsilon_0}\frac{1}{a^2}$$

and its direction is the same as that of $\mathbf{F}_3$. This latter feature is evident from the figure and the fact that the magnitudes of $\mathbf{F}_1$ and $\mathbf{F}_2$ are equal.

Example 19-5 A particle of charge q is on the perpendicular bisector and at a distance a away from a line charge of length $2l$ and of charge per unit length λ. Calculate the force on the particle.

Solution Consider this situation as shown in Figure 19-17. Let dF_x and dF_y represent the horizontal and vertical components of the force on the particle due to the charge $\lambda\, dx$ in an element of length dx at a distance x away from the y-axis. According to Coulomb's law,

$$dF_x = \frac{-q}{4\pi\epsilon_0}\lambda\, dx \frac{1}{a^2+x^2}\sin\theta$$
$$dF_y = \frac{q}{4\pi\epsilon_0}\lambda\, dx \frac{1}{a^2+x^2}\cos\theta \tag{19-7}$$

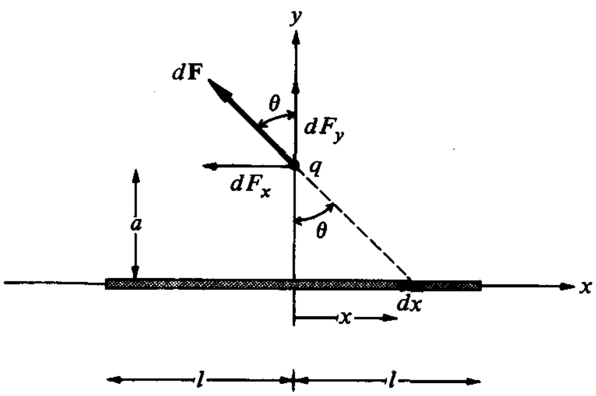


Figure 19-17

where the angle θ is as defined in the figure and the minus sign in the first equation reflects the fact that $d\mathbf{F}$ has a negative component along the x-axis. Reference to the figure shows that

$$\cos\theta = \frac{a}{(a^2+x^2)^{1/2}} \qquad \sin\theta = \frac{x}{(a^2+x^2)^{1/2}}$$

and thus (19-7) may be reexpressed in the form

$$dF_x = -\frac{\lambda q}{4\pi\epsilon_0}\frac{x}{(a^2+x^2)^{3/2}}\,dx$$
$$dF_y = \frac{\lambda q}{4\pi\epsilon_0}\frac{a}{(x^2+a^2)^{3/2}}\,dx \tag{19-8}$$

The total force $\mathbf{F}$ and its two components F_x and F_y are obtained by adding together contributions of the form in (19-8) for all elements of the line charge. Expressing this in the form of an integral, we find that

$$F_x = \int_{-l}^{l} dF_x = -\frac{\lambda q}{4\pi\epsilon_0}\int_{-l}^{l}\frac{x}{(a^2+x^2)^{3/2}}\,dx$$
$$F_y = \int_{-l}^{l} dF_y = \frac{\lambda q a}{4\pi\epsilon_0}\int_{-l}^{l}\frac{dx}{(a^2+x^2)^{3/2}} \tag{19-9}$$

where all constants have been taken out from under the integral signs. Reference to a

table of integrals shows that

$$\int \frac{x}{(x^2+a^2)^{3/2}}\,dx = -\frac{1}{(a^2+x^2)^{1/2}}$$

$$\int \frac{dx}{(x^2+a^2)^{3/2}} = \frac{x}{a^2}\frac{1}{(a^2+x^2)^{1/2}}$$

and making use of these values, the integrals in (19-9) are readily evaluated as follows:

$$F_x = -\frac{\lambda q}{4\pi\epsilon_0}\left[-\frac{1}{(x^2+a^2)^{1/2}}\right]\Bigg|_{-l}^{l} = 0$$

$$\begin{aligned} F_y &= \frac{\lambda q a}{4\pi\epsilon_0}\frac{1}{a^2}\left[\frac{x}{(x^2+a^2)^{1/2}}\right]\Bigg|_{-l}^{l} \\ &= \frac{\lambda q}{4\pi\epsilon_0 a}\left[\frac{l}{(a^2+l^2)^{1/2}} - \frac{-l}{(a^2+l^2)^{1/2}}\right] \\ &= \frac{2\lambda q l}{4\pi\epsilon_0 a(a^2+l^2)^{1/2}} \end{aligned}$$

The fact that $F_x = 0$ follows also by symmetry, since the x-component of the force due to the charge at the distance x cancels out that contribution due to the charge at $-x$. Or, equivalently, $F_x = 0$ follows since the integrand in the first equation of (19-9) is an odd function.

19-8 Limitations on charge

In the above discussion of electric charge and Coulomb's law, we have implicitly assumed that the electric charge q on a conductor can assume arbitrary positive and negative values. This is not correct! There is both an upper and a lower limit on the magnitude of the net amount of charge that can exist on a conductor.

The impossibility of having a very large amount of charge on a conductor is easy to see. The strong repulsive forces which would act between its components would produce severe stresses within the conductor and would ultimately cause it to explode. Normally this is not a problem, since, for a variety of practical reasons, it is not feasible to add enough charge to a rigid conductor to do this.

At the other end of the scale, we find an unambiguous lower limit to the smallest amount of charge that can be placed on a body. Experiment shows that there exists in nature a smallest subdivision of electric charge, and that all charge is an integral multiple of this smallest subdivision. We describe this by saying that electric charge is *quantized*, and refer to the smallest unit of charge as a *quantum of charge*. This quantum, for which the symbol e will be used, has the experimental value

$$e = 1.60219\times 10^{-19}\,\mathrm{C} \tag{19-10}$$

Note that no one in the laboratory or elsewhere has ever observed an amount of charge smaller in magnitude than this value. Nor has anyone ever observed an amount of charge that is not a positive or negative integral multiple of this quantum e. The charge of the electron (e), the negative pion (π^-), the negatively charged kaon (K^-), the omega-minus (Ω^-), the antiproton ($\bar{p}$), and all other negatively charged elementary particles have precisely the charge $-e$. Similarly, the charge on the proton (p), the positron (e^+), the positive pion (π^+), the positively charged kaon (K^+), the sigma-plus (Σ^+), and all other positively charged elementary particles have the precise value $+e$. No deviations from the values $\pm e$ or integral multiples of $\pm e$ have ever been observed.

Despite this fact that no one has ever observed a particle the magnitude of whose charge is less than e, certain theoretical proposals have been put forth, mainly by M. Gell-Mann, the 1969 Nobel Prize winner, according to which all observed particles are composites made up of certain fundamental particles whose charges are multiples of $\pm\frac{1}{3}e$. These particles have been given the name "*quarks*." According to these ideas, there are six distinct quarks (three ordinary ones, of charges $-\frac{1}{3}e$, $-\frac{1}{3}e$, and $+\frac{2}{3}e$, and three antiquarks, with the opposite signs of charge). For example, the positive pion (π^+) is a composite consisting of a quark of charge $+\frac{2}{3}e$ and an antiquark with charge $+\frac{1}{3}e$. By their very nature, quarks would be exceedingly difficult to detect, and as of this writing, and despite many efforts, they have not yet been observed. Until such time when they are, quarks must remain in the nature of a theoretical speculation, and for the present we can assume that the quantum of charge in (19-10) is the smallest subdivision of electric charge.

A question related to the existence of the quantum of charge is whether or not the positive quantum of charge has precisely the same magnitude as the negative one. That is, are the charges on the electron and proton exactly equal and opposite? A partial answer to this question has been previously given in Example 19-2, where it was shown that if the charge on an electron differed by only 1 part per million from the value $-e$, then extraordinarily large repulsive forces would exist between electrically isolated bodies. Since forces of such a large magnitude would be easily observable, from the fact that they have not been seen it follows that the charge on the electron and the proton must be very nearly, if not precisely, equal and opposite.

If it is assumed then that the charge on the electron is precisely equal to that on the proton, the results of many experiments may be summarized by a conservation law known as *the law of the conservation of charge.* According to this law, the algebraic sum of the electric charges involved in any process is conserved. If, for example, an atom is ionized, the initial electric charge is zero, as is the algebraic sum of the charge of the electron and ion $e+(-e)=0$ afterward. Or equally, if a certain positive electric charge is placed on a conductor, then an equal amount of negative charge must appear somewhere else. This principle of charge conservation, which has been

implicitly assumed to be valid in the above discussions, is very firmly established experimentally and no one has ever observed a violation of it under any circumstances.

†19-9 Limitations on the concept of a point charge

It has been noted previously that two small charged conducting spheres will satisfy Coulomb's law only if their separation distance is very large compared to their radii. For if they come close together, the charge distribution in each one will be influenced by that on the other, and the force between them will be modified accordingly. Nevertheless, it used to be believed not very long ago that for microscopic particles, such as electrons, protons, and nuclei, Coulomb's law would be applicable for arbitrarily small separation distances.

After the Stanford linear accelerator became operational in the 1950s it became possible for the first time to probe very small separation distances by use of energetic electrons. It was discovered by Robert Hofstadter and collaborators that the nucleus of an atom is not at all a point particle, as seemed to be implied by other experiments. But rather they found that the nucleus of each atom has an extended charge distribution; that is, that the positive charge on a nucleus is not concentrated at a point but is distributed more or less uniformly within a sphere whose radius is directly proportional to the cube root of the atomic number. Figure 19-18 shows the charge density—that is, the charge per unit volume—of several nuclei as a function of the distance r from their centers. Note that in the central region of the nucleus the charge density is constant but that it drops off very rapidly toward the outer edge of the nucleus.

It might be argued that since a heavy nucleus is actually a compound structure consisting of uncharged neutrons and positively charged protons,

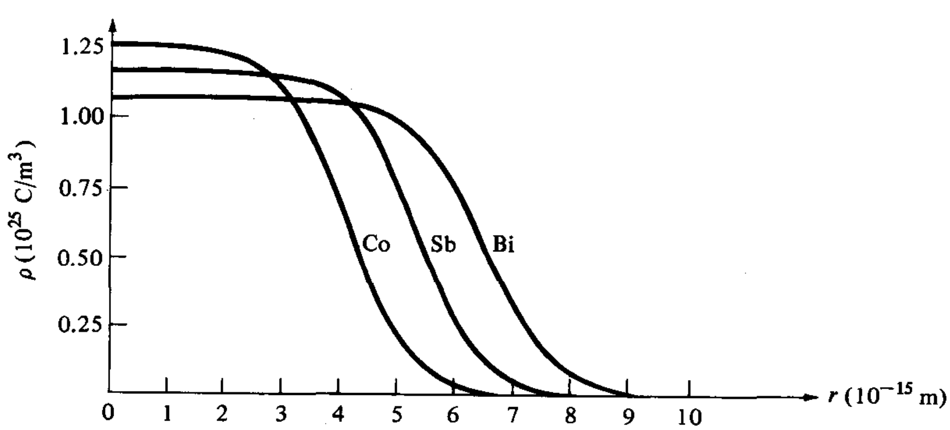


Figure 19-18 The charge density $\rho(r)$ of the nuclei Co, Sb, and Bi as a function of the radial distance r from the center. (Abstracted from the article by Robert Hofstadter, Ann. Rev. Nuc. Sci., 7, 231 (1957).)

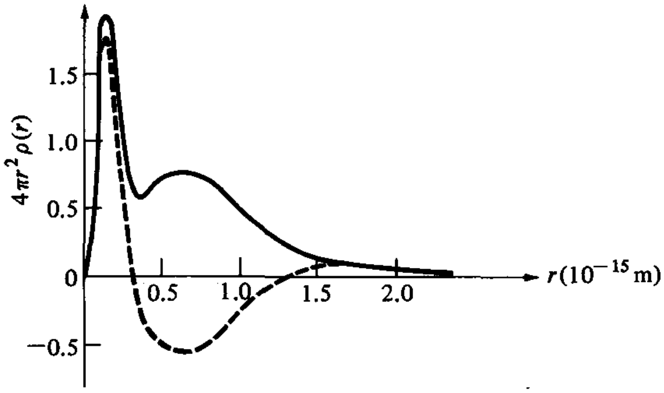


Figure 19-19 The weighted charge densities $4\pi r^2\rho(r)$ of the proton (solid curve) and the neutron (dashed curve) as a function of radial distance. The total area under the dashed curve is zero since the neutron is electrically neutral. (Abstracted from the article by D. N. Wilson, H. F. Shopper, and R. R. Wilson, Phys. Rev. Lett. 6, 286 (1961).)

results such as those in Figure 19-18 are not surprising. If on the other hand one were to probe, in the same way, the charge structure of a single proton or a neutron, then indeed they would be found to be *bona fide* point particles. Again, experiments carried out by Hofstadter demonstrated that even this is not true. But rather, as shown by the solid curve in Figure 19-19, the proton itself has a charge density $\rho(r)$, which extends over a distance $r \cong 10^{-15}$ meter. Also included on the same graph by a dashed curve is the charge density for the neutron. Even though the neutron is electrically neutral, it was discovered by scattering energetic electrons from deuterium that the neutron also has a charge structure. Specifically, for distances $r \leqslant 0.5 \times 10^{-15}$ meter, it was found to have a positive charge density, while for larger distances this charge density is negative by an appropriate amount so that the total charge Q_0

$$Q_0 = \int_0^\infty 4\pi r^2 dr \rho(r) \qquad \text{(neutron)}$$

vanishes.

At the present time it is generally believed that the electron, the muon, and the neutrino are point particles but that all other elementary particles have an extended structure of some type. It may well turn out that electrons and muons themselves also have an extended charge structure, but as yet no one has been able to suggest an experiment to resolve this very basic problem.

19-10 Summary of important formulas

The force **F** between two particles of charges q and Q separated by a distance r lies along the line joining the particles and has a magnitude given by Coulomb's law

$$F = \frac{1}{4\pi\epsilon_0}\frac{qQ}{r^2} \qquad \textbf{(19-3)}$$

The force **F** is repulsive if the charges have like sign and attractive otherwise.

For the case of more than two particles the force on any given one is the vector sum of the forces produced by each of the other particles.

QUESTIONS

1. A rubbed glass rod with its positive charge is brought near a spherical conductor suspended from an insulated string. It is observed that the sphere moves toward the rod, but that after making contact it jumps away and appears to be repelled by it. Explain these observations in terms of electron migrations within the conductor.
2. Consider the same situation as in Question 1, but suppose this time that the sphere is suspended by a conducting wire so that it is grounded. What is the behavior of the sphere now? Explain your answer and confirm your prediction by carrying out an experiment.
3. The essential components of an electroscope are a small, conducting sphere to which is attached a conducting wire, at the bottom of which hang two light strips of metallic foil (Figure 19-20). The strips are called leaves and are presumed to be far away from the sphere. It is observed as depicted in Figure 19-20b that if a rubbed glass rod is brought near the sphere, then the leaves of the electroscope diverge. Explain this effect in terms of electron motions.
4. In Figure 19-20b, what is the sign of the charge on the conducting sphere? What is the sign of the charge on the leaves?
5. A rubbed hard-rubber rod is brought near the conducting sphere of an electroscope. Explain what happens.

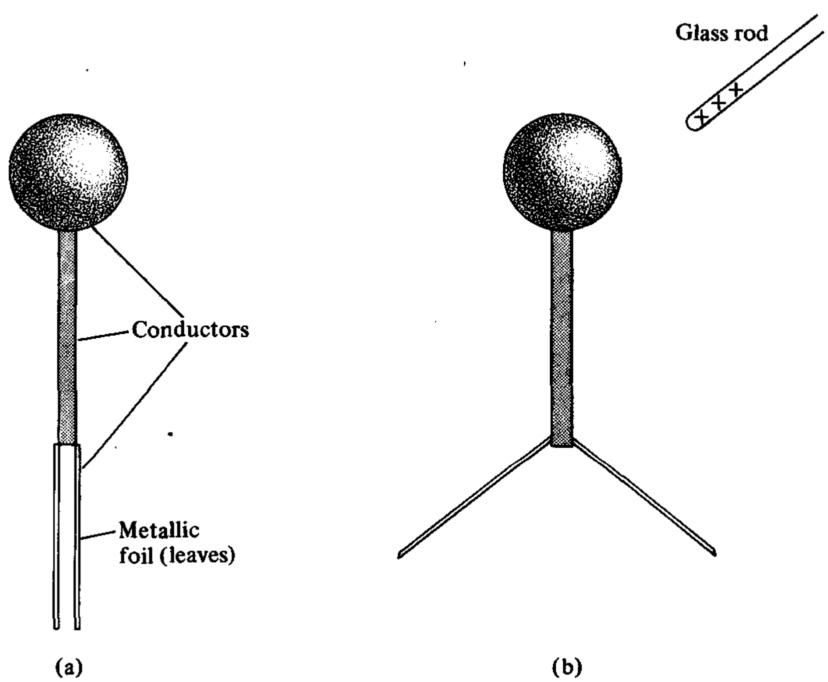


Figure 19-20

What is the sign of the charge on the sphere and on the leaves?

6. Consider again the situation in Figure 19-20b, but suppose that while the leaves are extended we ground the electroscope by touching the conducting sphere. Explain why the leaves now will collapse even though the rubbed glass rod is still near the sphere.
7. Subsequent to the collapse of the leaves in Question 6, suppose that the sphere is again disconnected from the ground. What will happen to the leaves? If the glass rod is removed, why will the leaves now diverge?
8. In Example 19-1 it was shown that the electric force between an electron and a proton is of the order of 10^{40} times greater than is their gravitational force. Explain why, despite this, it was necessary to include the weights of the two spheres in analyzing the setup in Figure 19-7.
9. Show that *if* the magnitudes of the charges on the electron and the proton differed by one part in 10^{18}, then the resultant electric force between two hydrogen atoms is comparable to their gravitational force. Why is this *not* a possible explanation for gravitational forces?
10. Suppose that an electroscope acquires an electric charge by the process outlined in Question 6. If the electroscope is allowed to remain in this state for some time, it is observed that the leaves collapse. What must have happened? Explain.
11. Explain why the leaves of a charged electroscope will collapse more readily when it is at a high elevation, say on the top of a mountain, than when it is at sea level. (*Note*: The first indications of the existence of cosmic rays came about as a result of this observation.)
12. In the construction of electroscopes, it is customary to surround the region below the sphere (see Figure 19-20a) by a conductor that is insulated from the metallic parts of the electroscope and is grounded. Can you think of a reason why this is desirable?
13. A particle of charge q is at a certain distance a away from a particle also of charge q. Where must a third charged particle be placed so that it experiences no electric force? Will this third particle exert a force on either of the others?
14. A very small insulating sphere is suspended from an insulating thread. Devise one experiment to determine whether the sphere is charged and one to ascertain the sign of this charge.
15. Repeat Question 14, but suppose this time that a *conducting* sphere is suspended from an insulating thread.
16. Present an argument that establishes that the net electric charge in the universe is "essentially" zero.
17. Two conducting spheres are found to attract each other with a certain electric force. Assuming that other charged bodies may be nearby, need the spheres necessarily have an electric charge? What about the case in which the spheres repel each other electrically?

PROBLEMS

In the following, use for the basic quantum of charge the value

$$\pm 1.6 \times 10^{-19}\ \mathrm{C}$$

and for the constant $1/4\pi\epsilon_0$ the value

$$9.0 \times 10^{9}\ \frac{\mathrm{N\text{-}m^2}}{\mathrm{C^2}}$$

In specifying a force, remember that both a magnitude and a direction are required.

1. (a) Calculate the amount of positive charge contained on all of the protons of 1 mole of monatomic hydrogen! (*Note*: This electric charge, equal to about 9.65×10^4 coulombs, is called the *faraday*. It represents the electric charge on 1 mole of any singly ionized substance.) (b) What is the amount of charge on the electrons in this gas?
2. Suppose that the protons and the electrons in 1 mole of monatomic hydrogen are separated and placed on opposite ends of the earth. What would be the force between them? (Assume that $R_E = 6.5 \times 10^6$ meters.)
3. Calculate the electric force between the two protons in a helium nucleus, assuming that their separation distance is 2×10^{-15} meter. Based on this result, what can you say about the strength of the nuclear forces between two protons when they are at this separation distance?
4. Two particles, each of charge 2 μC, are suspended at a separation distance of 30 cm by insulating strings each of length 0.5 meter (Figure 19-7). It is found that each of the strings make an angle of 30° with the vertical. Calculate (a) the mass of each particle, assuming that they are the same and (b) the tensions in the strings.
5. A nucleus of ^{3}H β-decays to ^{3}He according to the reaction

$$^3\text{H} \rightarrow {}^3\text{He} + e^- + \nu_e$$

where ν_e stands for the electrically neutral neutrino. If the separation distance between the electron and the ^{3}He nucleus immediately after the decay is 2×10^{-15} meter, calculate the force between them at this instant.
6. How much positive electric charge must be added to the sun and to the earth so as to nullify the gravitational force of attraction between them? Assume the following values: $M_S = 2.0 \times 10^{30}$ kg, $M_E = 6.0 \times 10^{24}$ kg, and $G = 6.67 \times 10^{-11}$ N-m^2/kg^2.
7. What would the charge on a proton have to be so that the electric and the gravitational forces between two protons cancel one another?
8. Two particles of charges q_1 and q_2 ($q_1 > q_2 > 0$) are separated by a certain distance d. Suppose that an amount of charge q is transferred from q_1 and q_2 so that the resultant charges are $(q_1 - q)$ and $(q_2 + q)$. For what value of q is the force of repulsion between the particles a maximum?
9. Show that if (x, y, z) are the coordinates of q in Figure 19-11, then the x-component of the force **F** is

$$F_x = \frac{qQ}{4\pi\epsilon_0} \frac{x}{[x^2 + y^2 + z^2]^{3/2}}$$

and find the corresponding forms for F_y and F_z.
10. Three particles, each of charge 3.0 μC, lie along a straight line at intervals of 0.5 meter. (a) Calculate the force on the middle particle. (b) Calculate the force on one of the end particles.
11. Particles having charges $-1.0\,\mu$C, $+2.0\,\mu$C, $+4.0\,\mu$C are located at the vertices of an equilateral triangle having sides 1 meter long. Calculate the magnitude of the force on (a) the $-1.0\,\mu$C particle; (b) the $+4.0\,\mu$C particle; and (c) the $+2.0\,\mu$C particle. (d) Show that the vector sum of these three forces is zero.
12. Two particles, of respective charges $-3.0\,\mu$C and $+4.0\,\mu$C, are at a separation distance of 0.3 meter. Where, along the line joining them, can a third charge be placed so that it experiences no electric force?
13. Two particles, each of charge $+q$, are placed on diagonally opposite vertices of a square of side a, and

two particles each of charge $-q$ are placed at the remaining vertices. Calculate (a) the force on one of the particles of charge q and (b) the force on one of the negatively charged particles.

14. A particle of charge q and mass m is attached to a spring of constant k and of equilibrium length l_0. Suppose that a second particle of charge $-q$ is brought within a distance a of the place where the spring is attached (Figure 19-21). Show that when equilibrium is reached, the spring has been stretched by an amount d, which satisfies the equation

$$\frac{q^2}{4\pi\epsilon_0}\frac{1}{(a-l_0-d)^2}=kd$$

What is the period of oscillations for small displacement from this equilibrium configuration? Assume that $a > l_0 + 3d$.

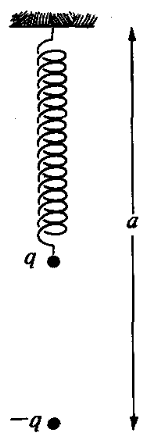


Figure 19-21

15. A particle of charge Q lies midway between two fixed identical charges of the same sign, each of magnitude q and separated by a distance $2b$.
(a) What is the force on Q?
(b) Suppose this particle is displaced from its original position by an amount y as shown in Figure 19-22. What is now the force acting on the particle?
(c) Show that if $y \ll b$, then if the particle is released from its position in Figure 19-22, it will oscillate with simple harmonic motion at angular frequency ω, given by

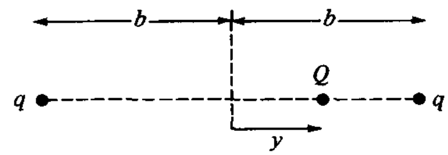


Figure 19-22

$$\omega=\left(\frac{qQ}{\pi\epsilon_0 m}\frac{1}{b^3}\right)^{1/2}$$

where m is the mass of the particle.

16. Repeat (b) and (c) of Problem 15, but this time assume that the particle is displaced along the perpendicular bisector of the line joining the two fixed particles. Assume also that $Qq < 0$.

17. Two particles of charges $+q$ and $-q$ are placed at the respective points $(a, 0, 0)$, $(-a, 0, 0)$ in a certain coordinate system. Show that the force on a charged particle of charge $Q(>0)$ located somewhere on the y-z plane (with coordinates $(0, y, z)$) points along the negative x-axis. What is the strength of this force as a function of y and z?

18. A particle of charge q lies along the axis of a uniform line charge of length l and of charge per unit length λ. Assume that the particle is at a distance a from the nearer end of the line charge as shown in Figure 19-23.

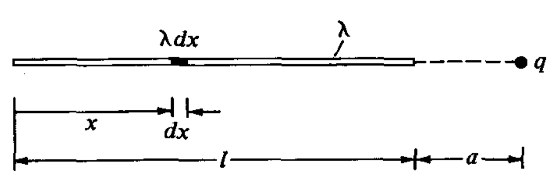


Figure 19-23

(a) Show that the force dF on the particle due to an element of

length dx is

$$dF = \frac{q\lambda\, dx}{4\pi\epsilon_0} \frac{1}{(a+l-x)^2}$$

(b) By integrating over all values of x from 0 to l calculate the total force on the particle.

*19. Eight particles, each of charge q, are distributed at relative angles of $2\pi/8 = \pi/4$ around a circle of radius a. A particle of charge Q is located on the axis of the circle and at a distance b from its center. Show that the magnitude of the force on Q is

$$\frac{2qQ}{\pi\epsilon_0} \frac{b}{(a^2+b^2)^{3/2}}$$

and find the direction of this force (*Hint*: Recall that force is a vector quantity and thus vector addition must always be used.)

20. Calculate the force on the particle in Figure 19-23, but assume now that λ is given by

$$\lambda = \lambda_0\left(1 - \frac{2x}{l}\right)$$

where λ_0 is a constant.

*21. A particle of charge Q is on the axis of a circular loop, of radius a and carrying a uniform charge per unit length λ.

(a) Calculate the horizontal and vertical components of the force $d\mathbf{F}$ due to an element of charge of length $a\,d\theta$ (Figure 19-24).

(b) Calculate the total force acting on Q.

*22. Calculate the magnitude of the force on the particle in Figure 19-24, but

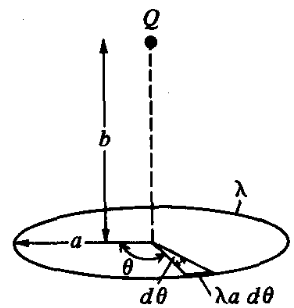


Figure 19-24

assume this time that the charge per unit length λ is

$$\lambda = \lambda_0 \sin\frac{\theta}{2}$$

where $\lambda_0 = $ constant.

20 The electrostatic field and Gauss' law

I have yet to see any problem, however complicated, which when you looked at it the right way did not become still more complicated.

P. ANDERSON

20-1 Introduction

Electrostatics is the study of charged particles at rest. Since the electric force between two charged particles is given directly by Coulomb's law, it might appear at first that with the statement of this law in Chapter 19 our study of electrostatics would be at an end. In actual fact nothing is farther from the truth. Coulomb's law, though basic to electrostatics, is only a first doorway to knowledge of this field.

Now the reason why Coulomb's law is only the beginning of electrostatics is due mainly to the fact that in the laboratory we are usually concerned with the behavior of electric charge in the presence of matter. Experiment shows that, in general, whenever electric charge is brought near matter, additional electric charges are induced in it. And since the distribution of this induced charge is not known *a priori*, it is not possible to apply Coulomb's law directly to these physically more interesting situations. If, for example, a charged particle is brought near an uncharged conducting body, the electric

charges—that is, the ions and electrons—in the conductor will experience electric forces which cause a redistribution of charge within the body. In order to be able to apply Coulomb's law to this case it is necessary to know precisely how this charge is distributed within the conductor. Only after this information has been obtained can Coulomb's law be used to calculate the force on the given particle.

Consider, for example, the problem of a particle of charge q (>0) near an electrically neutral but isolated conducting sphere (Figure 20-1). The presence of the positively charged particle causes some of the free electrons inside to migrate to the nearer side of the sphere and to leave a compensating positive charge on the opposite side. Now if the charge distribution on the sphere were known in detail, it would be possible, in principle, to calculate by use of Coulomb's law, the force between the sphere and the particle. We would multiply each element of charge in the sphere by $q/4\pi\epsilon_0$, divide by the square of the distance between the particle and the given element, and carry out a vector summation over all such charge elements. Unfortunately, the charge distribution on the sphere is not known and thus it is not possible to proceed this way.

Similarly, if a charged particle is brought near an insulator a redistribution of electric charge within the insulator also takes place. Again, until the nature of the resultant charge has been ascertained, we cannot calculate the force of attraction between the particle and the insulating body.

Now mainly for this purpose of describing in a quantitative way how electric charge in matter responds to an external charged body, it is convenient to introduce the concept of the *electric field.* As we shall see, not only is this physical construct of decisive utility in the solution of problems in electrostatics, but it also has a certain physical reality of its own. For the moment, let us focus attention on problems of electrostatics and think of the electric field simply as a convenient way of characterizing the distribution of charge in ordinary matter when it is subjected to an electric force. The more profound aspects of the electric field, and in particular its indisputable reality, will become apparent only after we have studied time-dependent phenomena.

20-2 The electric field

Consider, in Figure 20-2, a collection of particles of charges, $q_1, q_2, \ldots$. Let us bring up to these charges a *test particle* of (positive) charge q_0 and measure the electric force $\mathbf{F}$ that it experiences. In general, this force will be due to the charge on the given particles plus that induced in any matter that may be nearby. We define, provisionally, the *electric* or the *electrostatic field* $\mathbf{E}_0$ (due to the given charges) at the location of the test particle to be the force per unit charge on the test particle; that is,

$$\mathbf{E}_0 = \frac{1}{q_0}\mathbf{F} \qquad \textbf{(20-1)}$$

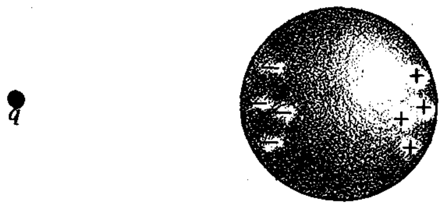


Figure 20-1

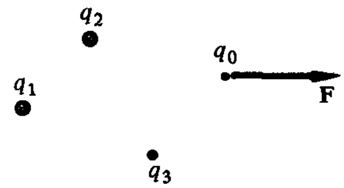


Figure 20-2

Now because of the fact that charge may be induced in matter, this provisional definition of the electric field is not completely satisfactory. The difficulty is that $\mathbf{E}_0$ depends not only on the given charge distribution but also on the strength of the test particle and its associated ability to induce charge in nearby matter. This means, for example, that $\mathbf{E}_0$ depends on the strength of the charge q_0 of the test particle and is not associated solely with the given collection of charged particles and the charge that these induce on nearby matter. Accordingly, we shall define the electric field $\mathbf{E}$ for the given charge distribution not by (20-1) but by its limiting form as the charge q_0 of the test particle becomes vanishingly small. Thus

$$\mathbf{E} = \lim_{q_0 \to 0} \frac{1}{q_0} \mathbf{F} \tag{20-2}$$

where $\mathbf{F}$ is the force on the test particle. The advantage of the definition in (20-2) is that the presence of the test particle with its very small charge will not cause any redistribution of charge in nearby matter. Thus the electric field as here defined is a property *only of the given charge distribution.* Note that, in general, the force $\mathbf{F}$ on the test particle varies with position, so the electric field $\mathbf{E}$ will also vary from point to point in space.

Consider a region in space in which there exists an electric field $\mathbf{E}$. If a particle of charge q is introduced into this region, it will experience the force $q\mathbf{E}$, provided that q is sufficiently small that no additional charges are induced in nearby matter. Otherwise the force on it will not be $q\mathbf{E}$ but rather $q\mathbf{E}'$, with $\mathbf{E}'$ the electric field due to the original charge plus that induced on nearby matter by the charge q on the particle. Unless a statement is made explicitly to the contrary, it will always be assumed that the charge q is sufficiently small that the force on it is $q\mathbf{E}$.

The unit of the electric field is that of force per unit charge. Hence the unit of $\mathbf{E}$ is the newton per coulomb (N/C).

Example 20-1 An electron of charge $q = -1.6 \times 10^{-19}$ coulomb is located in a region of space where there exists a uniform electric field of strength $E = 10^5$ N/C. Calculate:

(a) The force on the electron.
(b) Its acceleration.

Solution

(a) The force F on the electron is

$$F = qE = -1.6 \times 10^{-19}\,\mathrm{C} \times 10^5\,\mathrm{N/C}$$
$$= -1.6 \times 10^{-14}\,\mathrm{N}$$

The minus sign reflects the fact that this force is in a direction opposite to that of the electric field owing to the negative charge of the electron.

(b) According to Newton's second law, the acceleration a of the electron is F/m. Using the above value for F and the value $m = 9.1 \times 10^{-31}$ kg for the electron, we find that

$$a = \frac{1.6 \times 10^{-14}\ \text{N}}{9.1 \times 10^{-31}\ \text{kg}} = 1.8 \times 10^{16}\ \text{m/s}^2$$

This acceleration is directed along the force and is thus opposite to that of the electric field.

Example 20-2 Two particles of charges q and $-q$ are attached to a rigid rod of length $2b$, with the entire structure in a uniform electric field $\mathbf{E}$ (Figure 20-3). Calculate:

(a) The force on the structure.

(b) The torque about the center of the rod, assuming that it is at an angle θ with respect to the direction of the electric field.

Solution

(a) Reference to the figure shows that for $q > 0$, the force on the upper particle, $q\mathbf{E}$, points along the direction of the field and that on the lower particle, $-q\mathbf{E}$, is equal and opposite. Hence the net force on the structure is zero.

(b) To calculate the torque $\boldsymbol{\tau}$ on the structure about its center, recall from our studies in mechanics that the torque $\boldsymbol{\tau}$ produced by a force $\mathbf{F}$ about a given origin is

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$$

where $\mathbf{r}$ is a vector from the origin to the point at which the force acts. According to this definition, the force on the upper particle produces a torque $\boldsymbol{\tau}$ directed down into the plane of the diagram and with magnitude

$$\tau = bqE \sin\theta$$

since $|\mathbf{r}| = b$, $|\mathbf{F}| = qE$ and the angle between these vectors is θ. Applying the definition a second time but to the lower particle, we find precisely the same torque. Hence the torque $\boldsymbol{\tau}$ acting on the structure is directed perpendicularly down into the plane of Figure 20-3 and has the magnitude

$$\tau = 2bqE \sin\theta$$

20-3 The electric field of discrete particles

In this section and the next we make use of the definition in (20-2) to derive formulas for the electric field $\mathbf{E}$ associated with certain charge distributions.

Consider first the case of a single particle of charge q. To determine the electric field $\mathbf{E}$ at a *field point*, which is at a displacement $\mathbf{r}$ from the particle, let us bring up a small test particle of charge q_0 (>0) (Figure 20-4). According to Coulomb's law, the force $\mathbf{F}$ on the test particle is

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{qq_0}{r^2}\hat{\mathbf{r}}$$

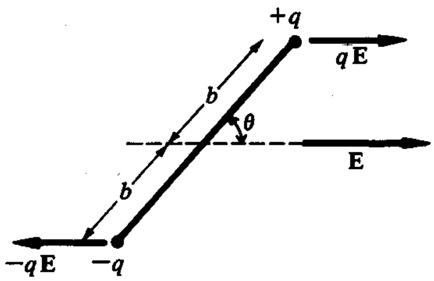


Figure 20-3

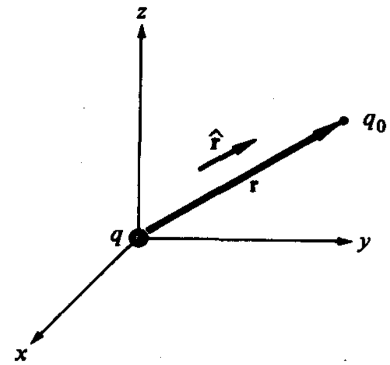


Figure 20-4

where $\hat{\mathbf{r}}$ is a unit vector along the direction from q to the test particle. Substitution into (20-2) yields

$$\begin{aligned}\mathbf{E} &= \lim_{q_0 \to 0} \frac{\mathbf{F}}{q_0} = \lim_{q_0 \to 0} \frac{1}{q_0} \frac{1}{4\pi\epsilon_0} \frac{qq_0}{r^2} \hat{\mathbf{r}} \\ &= \frac{q}{4\pi\epsilon_0} \frac{\hat{\mathbf{r}}}{r^2}\end{aligned} \qquad (20\text{-}3)$$

so for $q > 0$ the electric field is directed radially outward from the particle and varies inversely with the square of the distance from it. For a negatively charged particle the spatial variation is the same, but the field is directed radially inward toward the particle.

In a similar way we can obtain a formula for the electric field associated with more than one particle. Figure 20-5 shows a test particle of charge q_0 at a point that is at the distances r_1 and r_2 from two fixed particles of respective

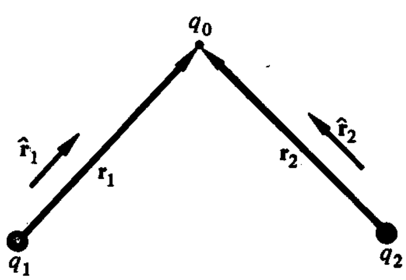


Figure 20-5

charges q_1 and q_2. Defining $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ to be the respective unit vectors from q_1 and q_2 to q_0, and making use of the superposition principle, we obtain

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} \frac{q_0 q_1}{r_1^2} \hat{\mathbf{r}}_1 + \frac{1}{4\pi\epsilon_0} \frac{q_0 q_2}{r_2^2} \hat{\mathbf{r}}_2$$

and thus, according to the definition in (20-2), the electric field $\mathbf{E}$ at the position of q_0 is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \left[\frac{q_1}{r_1^2} \hat{\mathbf{r}}_1 + \frac{q_2}{r_2^2} \hat{\mathbf{r}}_2 \right] \qquad (20\text{-}4)$$

The generalization of this formula to the case of N particles may be obtained by a straightforward extension of this argument. The result is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N} \frac{q_i \hat{\mathbf{r}}_i}{r_i^2} \tag{20-5}$$

where $\hat{\mathbf{r}}_i$ is a unit vector from the position of the ith particle of charge q_i to the field point, and $\{r_i\}$ are the distances of the particles from the field point.

It is interesting to note that (20-4) and its generalization to more than two particles may be interpreted by saying that the electric field at any given point due to a collection of charged particles is equal to the sum of the electric fields produced by each particle separately. This property, which is a direct consequence of the corresponding feature of Coulomb's law and (20-2), is usually described by saying that the *superposition principle* is also valid for the electric field. That is, the electric field due to a collection of particles is the sum of the fields produced by each particle separately.

Example 20-3 Two identical particles, each of charge q, are separated by a distance $2a$. Calculate the electric field at a point that lies:

(a) On the line joining the particles but is not between them.

(b) On the perpendicular bisector of the line joining them.

Solution Let us set up a coordinate system with the two particles located along the x-axis so that their coordinates are, respectively, $(a, 0)$ and $(-a, 0)$; see Figure 20-6.

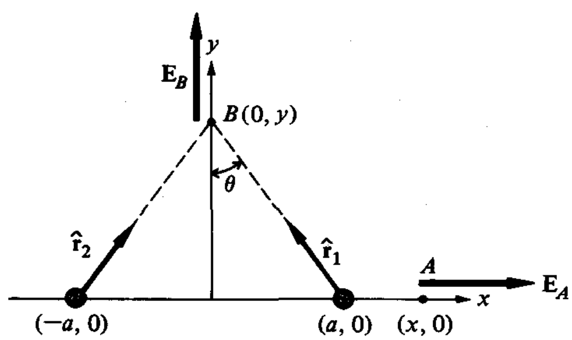


Figure 20-6

(a) If A is the field point with coordinates $(x, 0)$, then its distances from the two particles are $(x - a)$ and $(x + a)$, respectively. The unit vectors from each particle to the field point are parallel to the x-axis and thus to the unit vector $\mathbf{i}$ along this axis. Substituting these data into (20-5) and assuming that $x > a$, we obtain for the electric field $\mathbf{E}_A$ at point A

$$\mathbf{E}_A = \mathbf{i}\frac{q}{4\pi\epsilon_0}\left(\frac{1}{(x-a)^2} + \frac{1}{(x+a)^2}\right)$$

which as shown in the figure lies along the line joining the two particles.

(b) In a similar way, the electric field $\mathbf{E}_B$ at the point B with coordinates $(0, y)$ is

$$\begin{aligned}\mathbf{E}_B &= \frac{q}{4\pi\epsilon_0}\left(\hat{\mathbf{r}}_1 \frac{1}{y^2 + a^2} + \hat{\mathbf{r}}_2 \frac{1}{y^2 + a^2}\right) \\ &= \frac{q}{4\pi\epsilon_0} \frac{1}{y^2 + a^2}(\hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2)\end{aligned}$$

since the distance of each particle from the field point is $(a^2+y^2)^{1/2}$. The unit vectors $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ are as defined in the figure, and hence the sum $(\hat{\mathbf{r}}_1+\hat{\mathbf{r}}_2)$ is a vector parallel to the unit vector $\mathbf{j}$ along the y-axis. Since each of the unit vectors $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ has the component $\cos\theta$ along the y-axis, it follows that

$$\hat{\mathbf{r}}_1+\hat{\mathbf{r}}_2=2\mathbf{j}\cos\theta$$

Finally, reference to the figure shows that $\cos\theta = y/(y^2+a^2)^{1/2}$, so the final formula for $\mathbf{E}_B$ is

$$\mathbf{E}_B=\mathbf{j}\frac{q}{2\pi\epsilon_0}\frac{y}{(y^2+a^2)^{3/2}}$$

Thus, $\mathbf{E}_B$ lies along the perpendicular bisector and decreases as $1/y^2$ for large y.

20-4 The electric field for continuous charge distributions

Although all electric charge in matter is localized on very small particles, that is, on electrons and ions, for many applications to problems involving macroscopic bodies it is convenient to think of these charge distributions as varying in a continuous way. This means that the actual discrete charge distribution is replaced by a smoothed-out, continuous *charge density* that represents, on the average, the charge per unit volume at any point in the body under consideration. The calculation of the electric field associated with these charge distributions is very similar to that of the discrete case considered in Section 20-3. The basic difference is that the sum in (20-5) must now be replaced by an appropriate integral. Let us illustrate the method by reference to several examples.

Example 20-4 Calculate the electric field at a perpendicular distance a from an infinite line charge of charge per unit length λ.

Solution As in Figure 20-7, let us set up a coordinate system with the x-axis along the line charge and with the y-axis through the field point. The electric field $d\mathbf{E}$ at this

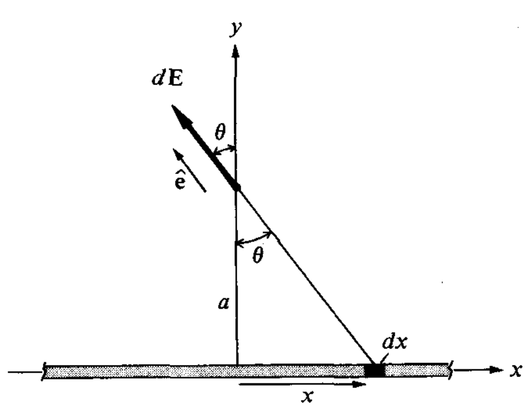


Figure 20-7

point due to the element of charge λdx at a distance x from the y-axis is

$$d\mathbf{E} = \frac{\lambda\, dx}{4\pi\epsilon_0} \frac{1}{(a^2+x^2)} \hat{\mathbf{e}}$$

with $\hat{\mathbf{e}}$ a unit vector parallel to the line from dx to the field point.

Now, by symmetry, the component of $d\mathbf{E}$ along the x-axis due to the charge element at x will be canceled out by a contribution to the *total* electric field by the charge element of strength $\lambda\, dx$ located at the point $-x$. Thus, the total electric field will have only a component along the y-axis and we need only consider the component of $d\mathbf{E}$ along this axis. Calling this component dE, we have

$$dE = \mathbf{j} \cdot d\mathbf{E} = \frac{\lambda\, dx}{4\pi\epsilon_0} \frac{1}{a^2+x^2} \cos\theta = \frac{\lambda\, dx}{4\pi\epsilon_0} \frac{a}{(a^2+x^2)^{3/2}}$$

where in the final equality we have used $\cos\theta = a/(a^2+x^2)^{1/2}$. The total electric field E is obtained by summing up these contributions for all values of x. Expressing this in the form of an integral, we have

$$E = \int dE = \frac{\lambda a}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{dx}{(a^2+x^2)^{3/2}}$$

where the constant $\lambda a/4\pi\epsilon_0$ has been taken out from under the integral sign.

To evaluate the integral let us make in the integrand the substitution $x = a\tan\theta$. The upper and lower limits of the integral become $\pi/2$ and $-\pi/2$, respectively, and the infinitesimal dx becomes $dx = a\sec^2\theta\, d\theta$. Making these changes we obtain

$$\begin{aligned} E &= \frac{\lambda a}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{dx}{(a^2+x^2)^{3/2}} = \frac{\lambda a}{4\pi\epsilon_0} \int_{-\pi/2}^{\pi/2} \frac{a\sec^2\theta\, d\theta}{[a^2+a^2\tan^2\theta]^{3/2}} \\ &= \frac{\lambda a}{4\pi\epsilon_0} \int_{-\pi/2}^{\pi/2} \frac{a\sec^2\theta\, d\theta}{a^3\sec^3\theta} = \frac{\lambda}{4\pi\epsilon_0 a} \int_{-\pi/2}^{\pi/2} \cos\theta\, d\theta \\ &= \frac{\lambda}{4\pi\epsilon_0 a} \sin\theta \Big|_{-\pi/2}^{\pi/2} = \frac{\lambda}{2\pi\epsilon_0 a} \end{aligned}$$

where the third equality follows by use of the identity $1+\tan^2\theta = \sec^2\theta$, and the fourth since $\cos\theta = 1/\sec\theta$.

Thus we conclude that the field at a distance a from an infinite line charge of charge per unit length λ is perpendicular to the line charge and has a strength E given by

$$E = \frac{\lambda}{2\pi\epsilon_0 a} \tag{20-6}$$

Example 20-5 A thin, circular loop of radius a has a charge per unit length λ. Calculate the electric field at a point on the axis of the loop at a distance b from it.

Solution Consider, in Figure 20-8a, the electric field $d\mathbf{E}$ due to an element of charge of strength $\lambda a\, d\theta$ and which lies at an angular separation θ from a fixed reference line. According to (20-3),

$$d\mathbf{E} = \frac{\lambda a\, d\theta}{4\pi\epsilon_0} \frac{\hat{\mathbf{e}}}{a^2+b^2}$$

where ê is a unit vector oriented as shown. Now as indicated in Figure 20-8b, the component of $d\mathbf{E}$ parallel to the plane of the loop and due to the charge element located at θ is equal and opposite to the corresponding component produced by the charge element at $(\pi + \theta)$. Thus, as in Example 20-4, the electric field due to the entire loop will have a nonvanishing component only along the axis of the loop and only the component dE of the vector $d\mathbf{E}$ along this direction is of interest. Since the component of the unit vector ê along this direction is $\cos\phi$ and since $\cos\phi = b/(a^2 + b^2)^{1/2}$, it follows that

$$dE = \frac{\lambda a\, d\theta}{4\pi\epsilon_0} \frac{b}{(a^2 + b^2)^{3/2}}$$

Integrating this formula over all values of θ from 0 to 2π, we obtain for the total electric field E

$$E = \int dE = \int_0^{2\pi} \frac{\lambda a\, d\theta}{4\pi\epsilon_0} \frac{b}{(a^2 + b^2)^{3/2}} = \frac{\lambda ab}{4\pi\epsilon_0 (a^2 + b^2)^{3/2}} \int_0^{2\pi} d\theta$$
$$= \frac{\lambda ab}{4\pi\epsilon_0 (a^2 + b^2)^{3/2}} 2\pi$$

Thus the electric field is directed along the axis of the loop and has the magnitude

$$E = \frac{\lambda ab}{2\epsilon_0 (a^2 + b^2)^{3/2}} \tag{20-7}$$

Example 20-6 A nonconducting circular disk of radius R has a charge per unit area σ. Calculate the electric field at a point on the axis and at a distance b from the disk (Figure 20-9).

Solution Consider, as shown in the figure, an infinitesimal circular ring of the disk of radius r and of thickness dr. The area of this ring is $2\pi r\, dr$, and, since the charge per unit area on the disk is σ, the total charge on the ring is $\sigma 2\pi r dr$. According to (20-7), the electric field dE due to the charged ring in Figure 20-9 is

$$dE = \frac{rb\sigma\, dr}{2\epsilon_0 (r^2 + b^2)^{3/2}}$$

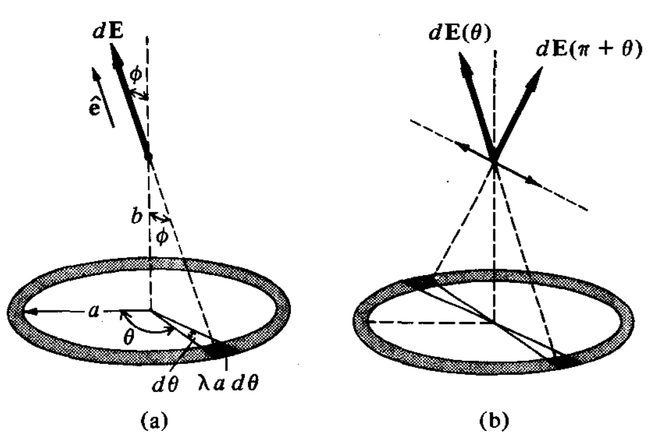


Figure 20-8

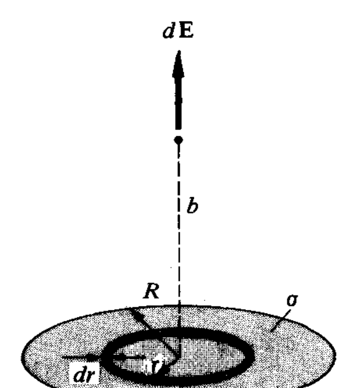


Figure 20-9

since the radius of the ring is r and its charge per unit length is now $2\pi\sigma r\,dr/2\pi r = \sigma\,dr$. As shown in the figure, this contribution to the electric field lies along the axis of the disk. The total electric field **E** also lies along this axis and its magnitude is obtained by summing up the contributions due to rings with radii from 0 to R. Expressing this sum as an integral we find that

$$E = \int dE = \int_0^R \frac{\sigma b}{2\epsilon_0} \frac{r\,dr}{(r^2+b^2)^{3/2}}$$
$$= \frac{\sigma b}{2\epsilon_0}\int_0^R \frac{r\,dr}{(r^2+b^2)^{3/2}} = \frac{\sigma b}{2\epsilon_0}\left[-\frac{1}{(b^2+r^2)^{1/2}}\right]\Bigg|_0^R$$
$$= \frac{\sigma b}{2\epsilon_0}\left[\frac{1}{b} - \frac{1}{(R^2+b^2)^{1/2}}\right]$$

so that the electric field E due to the entire disk is directed along the axis of the disk and has the magnitude

$$E = \frac{\sigma}{2\epsilon_0}\left[1 - \frac{b}{(R^2+b^2)^{1/2}}\right] \tag{20-8}$$

For the special case of an infinite plane of charge—that is, for $R \gg b$—the relation in (20-8) reduces to

$$E = \frac{\sigma}{2\epsilon_0} \tag{20-9}$$

and we note that this is independent of the distance b from the plane. In other words, the electric field due to an infinite nonconducting plane of charge is directed perpendicular to—and away from—the plane and has the *constant value* $\sigma/2\epsilon_0$ everywhere.

20-5 Geometrical representation of the electric field

Given any distribution of charge it is convenient to represent its associated electric field by a series of *directed lines*, which are called *electric field lines.* These lines are constructed so that they have the following two properties:

1. At each point along a given line, the tangent to the line (directed in the same sense as is the line) is parallel to the electric field at that point.
2. The number of electric field lines in any region of space is proportional to the strength of the electric field in that region.

Thus, as we go from a region where the electric field is strong—say in the neighborhood of a charged particle—to a region of relatively weaker electric field, the density of field lines decreases. Figure 20-10a shows a planar view of the electric field lines in the neighborhood of a positively charged particle. According to (20-3), the electric field is directed radially outward and decreases inversely with the square of distance. As evidenced by the arrowheads, the field lines in Figure 20-10a are also directed radially outward

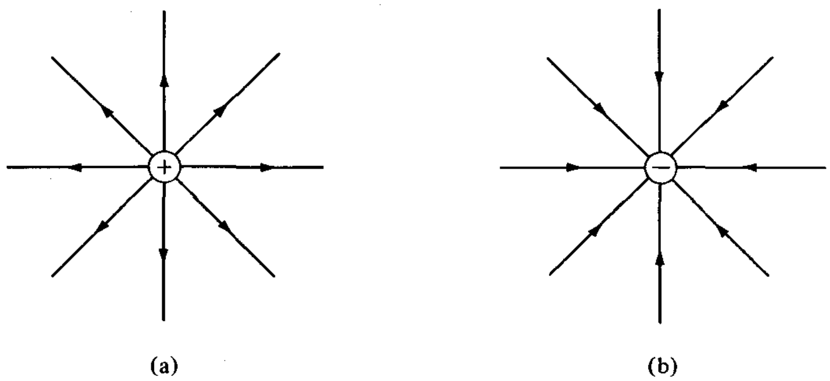


Figure 20-10

and their density decreases as we recede from the particle. Figure 20-10b shows the field lines for a negative particle. Note that here the field lines are oriented radially inward but that again in accordance with (20-3), the density of these lines increases as we approach the particle.

Because of the inverse square nature of Coulomb's law, it follows that in the immediate neighborhood of a charged particle the electric field is essentially determined by that of the particle alone. Thus, regardless of what other matter may be present, the electric field lines very close to a charged particle must always have the form in Figure 20-10. But what about the field lines in regions of space where the electric field is not dominated by a single particle? The following two properties of electric field lines are fundamental in this connection and are useful not only as an aid in visualizing field lines but are also indispensable in their construction. These properties are:

1. No two field lines can ever cross each other except at a point where there exists a charged particle.
2. All field lines are continuous in all regions of space containing no electric charge. Thus a field line must originate on a positively charged particle and terminate on a negatively charged particle, but no field line can ever originate or terminate at any point where there is no electric charge.

According to these properties, for example, the field lines near a charged body *cannot* have the structure depicted in Figure 20-11. The lines A, B, and C do not represent possible field lines since they either cross other field lines or else originate or terminate at a point at which there is no electric charge.

To justify the first of the above two properties, that in a charge-free region two field lines cannot cross, we may argue as follows. Suppose two field lines did cross. Then at such a point of crossing there would *not* be a unique tangent. This in turn would contradict the defining property of field lines that the tangent to a field line at a given point must be directed along the electric field at that point.

With regard to the second property, namely that all field lines in empty space must be continuous, it is unfortunately *not* possible to give a clean

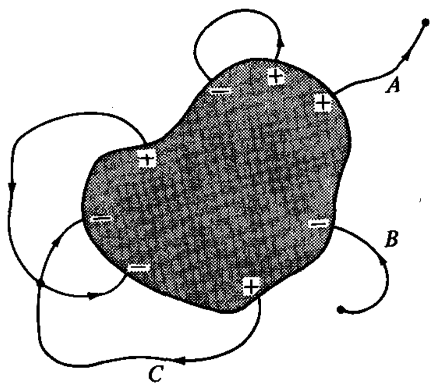


Figure 20-11

argument at an elementary level. Suffice it to say that in deriving this very crucial property of the continuity of the field lines, the fact that the electric field varies precisely with the inverse square of the distance from the particle plays a decisive role. If, for example, the exponent "2" in the inverse-square Coulomb law differed by any small amount from its established value of 2, then indeed electric field lines would *not* be continuous.

To illustrate the matter, consider, in Figure 20-12, three spherical surfaces concentric with a charged particle. Because of the continuity of electric field lines it is necessary that the same number of these lines cross the surface of each of the three spheres. For if this were not so, it would follow that at least one field line must originate or terminate in the region between the spheres and this would violate the continuity property of electric field lines in a charge-free region. (See Example 20-8 for a more detailed discussion of this matter.)

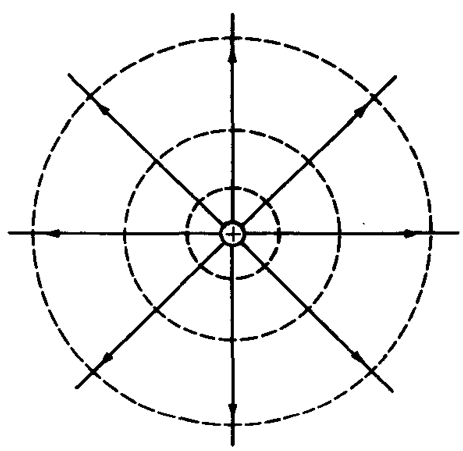


Figure 20-12

Figure 20-13 shows a planar view of the field lines associated with certain charge configurations. Parts (a) and (b) portray the field lines associated with two particles of opposite and of the same charge, respectively. The particular charge configuration shown in Figure 20-13a of equal and opposite

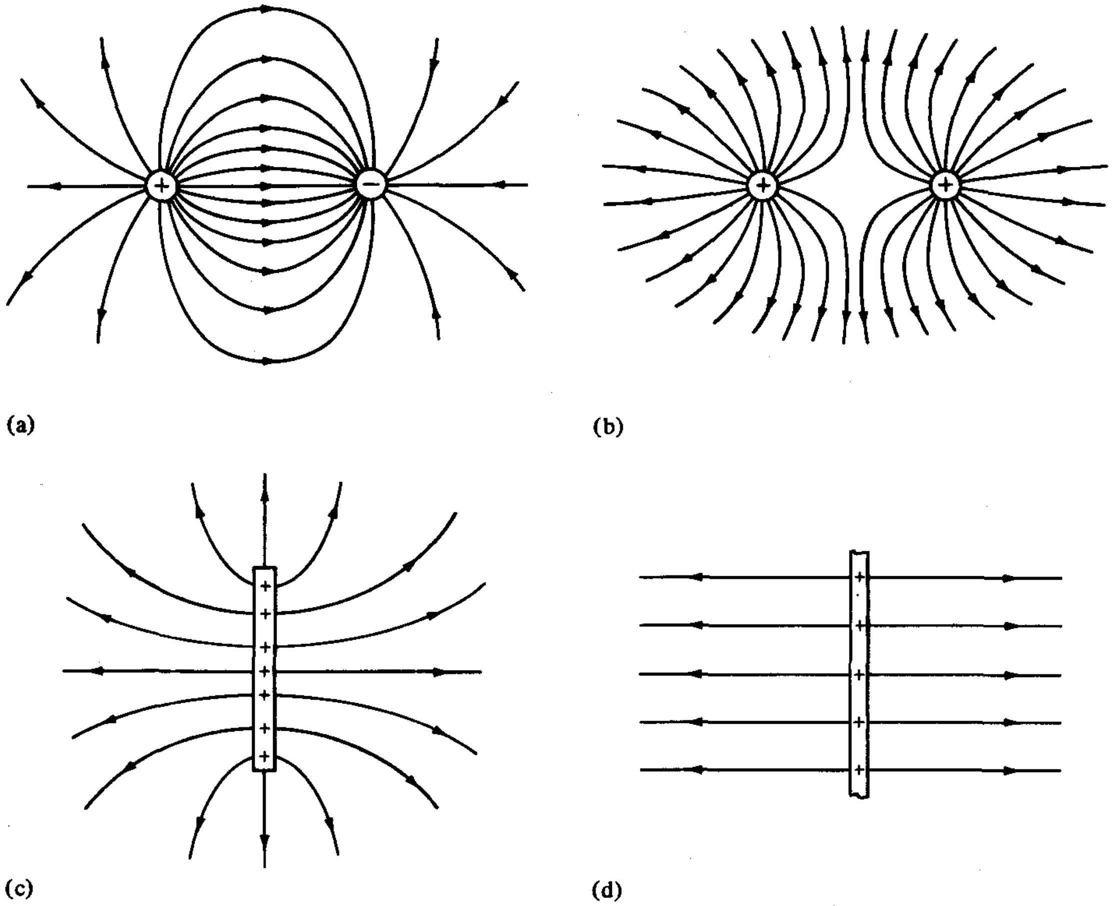


Figure 20-13

charges, q and $-q$, is known as an *electric dipole*, and the associated field as a *dipole field*. Note in each case that the field lines originate on the positive charge and terminate on the negative charge. For the case of the two positive charges, the endpoints of all field lines could not be included in the figure, and these lines can be thought of as terminating on negative charge at "infinity." Figure 20-13c shows the field lines associated with an edge-on view of a uniformly charged, nonconducting disk. Consistent with (20-8), the field on the axis of the disk is parallel to the axis and decreases away from the surface of the disk. Finally, Figure 20-13d shows the field lines associated with an infinite plane of charge. Consistent with (20-9), the field lines are uniform and perpendicular to the plane. Note that each of the diagrams in Figure 20-13 should be viewed in a three-dimensional perspective. The totality of field lines associated with the *dipole* in Figure 20-13a, for example, is obtained by rotating the figure about the line joining the two particles.

20-6 Electric flux

Consider, in Figure 20-14, a certain closed surface S. Let ΔS represent an infinitesimal element of area of S and define an associated vectorial area

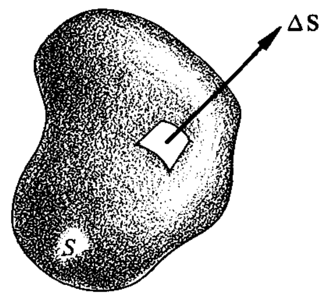


Figure 20-14

element **ΔS** to be a vector that has the magnitude of the area element ΔS itself and is directed perpendicularly outward from the surface. If, for example, the surface S is a sphere, then each of its area elements **ΔS** is directed radially outward from the center of the sphere.

Consider now an area element **ΔS** associated with a given surface S in a region where there exists also an electric field **E**. We define the *flux* $\Delta\Phi$ of the electric field **E** through this area by the formula

$$\Delta\Phi = \mathbf{E} \cdot \Delta\mathbf{S} \tag{20-10}$$

that is, by the dot product of the electric field and the area element **ΔS**. If θ is the angle between **E** and **ΔS**, then an equivalent way of expressing this definition is

$$\Delta\Phi = E\,\Delta S \cos\theta \tag{20-11}$$

As shown in Figure 20-15, the electric field flux through **ΔS** can be either positive, negative, or zero. The third possibility occurs if the electric field line lies in the surface element itself so that $\theta = \pi/2$.

More generally, consider a surface S composed of area elements **ΔS** to each of which a unique sense for the outward normal has been assigned. The total flux Φ_S through S is a sum of contributions of the type in (20-10); that is,

$$\Phi_S = \sum \mathbf{E} \cdot \Delta\mathbf{S} \tag{20-12}$$

where the sum extends over all area elements comprising the given surface

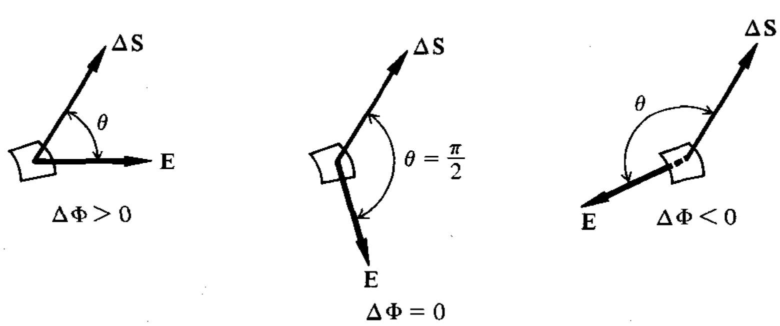


Figure 20-15

S. In the limit as these area elements tend to zero, the sum in (20-12) is defined to be the "surface integral"

$$\Phi_S = \int_S \mathbf{E} \cdot d\mathbf{S} \tag{20-13}$$

which may be viewed as simply another way of writing the limit of the sum in (20-12). For the special case that the surface S is closed and $d\mathbf{S}$ is directed outward, it is customary to write (20-13) in the form

$$\Phi_S = \oint_S \mathbf{E} \cdot d\mathbf{S} \qquad \text{(closed surface } S\text{)} \tag{20-14}$$

where the circle superimposed on the integral sign serves as a reminder that the surface S over which the integral is to be carried out is closed. Except for the very special cases to be considered below we shall *not* be concerned with evaluating surface integrals directly. However, we do need to give a meaning to these integrals and for this purpose it suffices to think of them as the limit of the sum in (20-12). From a physical point of view, the electric flux Φ_S is a measure of the number of field lines that go through the given surface.

Example 20-7 Calculate the flux through a planar surface of area A, which lies near an infinite, nonconducting plane carrying a charge per unit area σ. Assume that, as in Figure 20-16, α is the angle between the two planes and that $\hat{\mathbf{n}}$ is a unit vector normal to the surface.

Solution According to (20-9), the electric field is constant and normal to the charged plane and has magnitude $\sigma/2\epsilon_0$. Thus, since the angle between $\mathbf{E}$ and $\Delta\mathbf{S}$ is α, it follows from (20-11) that

$$\Phi = E\,\Delta S \cos\theta = \frac{\sigma}{2\epsilon_0} A \cos\alpha$$

Example 20-8 Calculate the flux through a spherical surface of radius R due to a particle of charge q at its center (Figure 20-17).

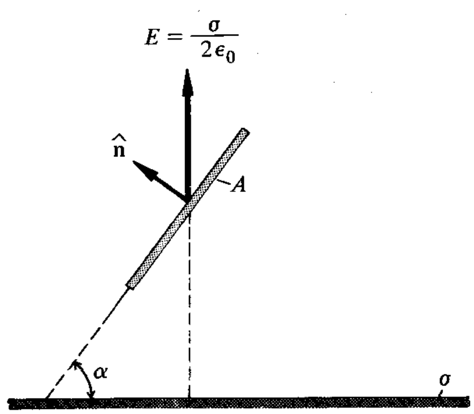


Figure 20-16

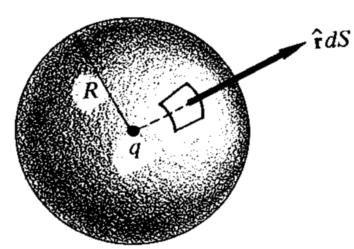


Figure 20-17

Solution According to (20-3), the electric field $\mathbf{E}$ at the distance R is

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0}\frac{1}{R^2}\hat{\mathbf{r}}$$

where $\hat{\mathbf{r}}$ is a unit vector directed along a radius. For the sphere, the normal to its surface is along the radius and thus

$$d\mathbf{S} = \hat{\mathbf{r}}\, dS$$

By use of these formulas, the integrand in (20-14) becomes

$$\begin{aligned} \mathbf{E}\cdot d\mathbf{S} &= \frac{q}{4\pi\epsilon_0}\frac{1}{R^2}\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}\, dS \\ &= \frac{q}{4\pi\epsilon_0}\frac{1}{R^2}\, dS \end{aligned}$$

where in the second equality we have used the fact that $\hat{\mathbf{r}}\cdot\hat{\mathbf{r}} = 1$, which is valid for any unit vector. Furthermore, since the quantity $q/4\pi\epsilon_0 R^2$ is constant everywhere on the surface of the sphere it follows that Φ_S in (20-14) may be written

$$\begin{aligned} \Phi_S &= \oint_S \mathbf{E}\cdot d\mathbf{S} = \frac{q}{4\pi\epsilon_0}\frac{1}{R^2}\oint dS = \frac{q}{4\pi\epsilon_0}\frac{1}{R^2}4\pi R^2 \\ &= \frac{q}{\epsilon_0} \end{aligned}$$

where the third equality follows since for a sphere $\oint dS = 4\pi R^2$. Thus we have in this case

$$\Phi_S = \frac{q}{\epsilon_0}$$

which is independent of the radius of the sphere! In the next section we shall see that this result is true not only for a spherical surface but for *any* closed surface.

20-7 Gauss' law

The notion of electric flux enables us to state another very important property of the electric field. This property is known as *Gauss' law*, and is intimately related to the continuity property of electric field lines. It states:

> *The total flux out of* any *closed surface S is the product of* $1/\epsilon_0$ *and the algebraic sum of* all *the charges inside S.*

In mathematical terms, Gauss' law is

$$\oint_S \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0}Q \qquad \textbf{(20-15)}$$

where Q is the net charge, due allowance being made for sign, inside S and where in evaluating the surface integral each of the area elements $d\mathbf{S}$ is directed perpendicularly outward. Note that (20-15) is valid only for a *closed* surface S, but is applicable regardless of its shape. Thus in Figure 20-18 the flux out of the closed surfaces S_1 and S_2 are both zero since the total charge inside each one vanishes. Correspondingly, the flux out of S_3 is $+q/\epsilon_0$ and that out of S_4 is $-q/\epsilon_0$. We shall refer to the closed surface S implied in the integral in (20-15) as a *Gaussian surface.* Thus S_1, S_2, S_3, and S_4 are all Gaussian surfaces.

By making use of the concepts of the vector calculus, it is possible to derive Gauss' law by starting with the definition for the electric field in (20-2). Unfortunately, these methods are largely beyond our means. Suffice it to say that just as for the continuity property of the electric field lines, the fact that the Coulomb force law varies precisely with the inverse square of the distance plays a decisive role in such a proof. If the exponent in Coulomb's law were not precisely equal to 2, then Gauss' law would *not* be valid.

To confirm this fact that Gauss' law is valid only for an inverse-square force, suppose that this exponent were actually $(2+\eta)$, where η is some nonzero number. Then the "electric field" $\boldsymbol{\epsilon}$ associated with a particle of charge q would be

$$\boldsymbol{\epsilon} = \frac{q}{4\pi\epsilon_0}\frac{1}{r^{2+\eta}}\hat{\mathbf{r}}$$

with $\hat{\mathbf{r}}$ a unit vector directed radially outward from the particle. This time if we calculate the flux Φ_S through a spherical Gaussian surface of radius R and centered at the particle, then following the same steps as in Example 20-8, we find that

$$\Phi_S = \oint_S \boldsymbol{\epsilon}\cdot d\mathbf{S} = \frac{q}{4\pi\epsilon_0}\oint \frac{dS}{R^{2+\eta}} = \frac{q}{4\pi\epsilon_0}\frac{1}{R^{2+\eta}}4\pi R^2$$
$$= \frac{q}{\epsilon_0}\frac{1}{R^{\eta}}$$

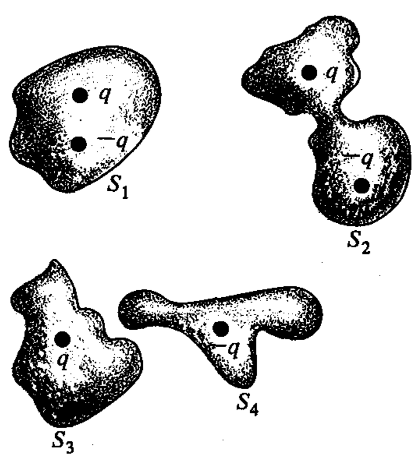


Figure 20-18

This varies with the radius R of the sphere, and hence the outward flux depends not only on the charge inside the surface but on the details of the surface as well. Thus, Gauss' law is *not* valid in this case.

Gauss' law can also be made plausible for the case where there is no charge inside the Gaussian surface. Consider, for example, the Gaussian surface and the field lines in Figure 20-19. Since, by hypothesis, there is no charge inside, it follows from the continuity property of field lines that every field line that enters the surface, such as at A, must leave the surface at some other point, such as B. Moreover, at A the angle θ between the electric field and the outward normal is obtuse and thus, according to (20-11), the flux at this point is negative. Correspondingly, at the point B the flux is positive, so it is at least possible for a cancellation to take place. Equivalently, we may argue that the fact that the total flux out of a closed surface containing no charge vanishes is due to the fact that no field lines can be created in its interior; thus every field line that enters the surface must leave it at some other point. If, on the other hand, some positive charge is inside the surface, then additional field lines will originate here and the flux would be positive. Correspondingly, if there is an excess of negative charge inside the surface, then additional field lines will enter it and the flux out of the surface will be negative.

Example 20-9 Consider the Gaussian surface *ABCDFGHJ* in Figure 20-20. Assuming that there is no charge in its interior, confirm the fact that the flux out of this surface vanishes. Assume the existence of a uniform electric field **E** perpendicular to face *ABCD*.

Solution Because of the given direction of the electric field, the only contribution to the flux integral in (20-14) comes from the two faces *ABCD* and *FGHJ*. The outward normals of the remaining four faces are always perpendicular to the electric field and thus there is no contribution to the flux along these.

The flux Φ_1 through the lower face *ABCD* is

$$\Phi_1 = -ES$$

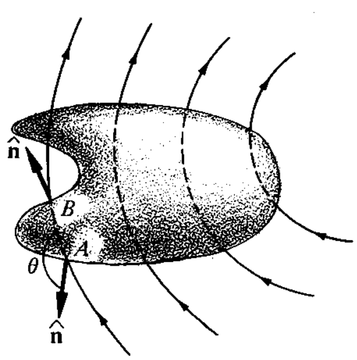


Figure 20-19

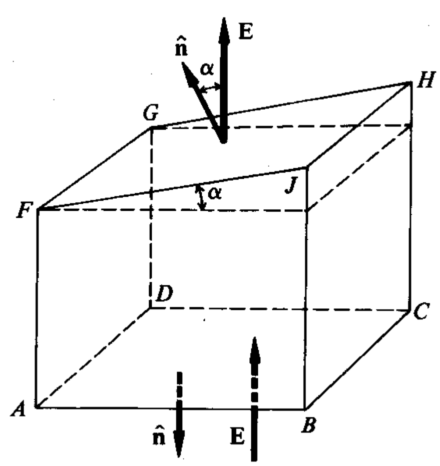


Figure 20-20

where S is the area of this face and the minus sign is due to the fact that the outward normal $\hat{\mathbf{n}}$ to this face is directed opposite to $\mathbf{E}$. Thus the angle θ in (20-11) is π and since $\cos \pi = -1$, the negative sign follows.

On the upper surface, $FGHJ$, the angle between the unit normal $\hat{\mathbf{n}}$ and $\mathbf{E}$ is α. According to (20-11), the flux Φ_2 is thus

$$\Phi_2 = E \cos \alpha \, S_2$$

where S_2 is the area of this upper surface. It is apparent from the figure that S_2 is related to the area S of the lower face by

$$S_2 = \frac{S}{\cos \alpha}$$

since, for example, the lengths AB and FJ are related by $FJ = AB/\cos \alpha$. Substituting this value for S_2 into the formula for Φ_2, the factor $\cos \alpha$ cancels and we obtain

$$\Phi_2 = ES$$

Finally, combining this formula with the formula for Φ_1 we conclude that, consistent with Gauss' law, the total outward flux $(\Phi_1 + \Phi_2)$ through the surface vanishes.

20-8 The electric field in the presence of a conductor

If a charged body, such as the glass rod in Figure 20-21, is brought near a conductor, induced charge will appear on it. The resultant total electric field $\mathbf{E}$ is due to both charges: the original charge on the rod plus that induced on the conductor. Because of the fact that the strength and the location of the induced charge are not generally known, we cannot make use of the methods of Sections 20-3 and 20-4 to calculate the total field $\mathbf{E}$. In this section we shall show that by use of Gauss' law it is possible, nevertheless, to establish the validity of the following properties concerning the distribution of induced charge on a conductor and its effect on $\mathbf{E}$.

1. $\mathbf{E} = 0$ everywhere *inside* the conductor.
2. There is no net charge *inside* the conductor.
3. $\mathbf{E}$ is everywhere perpendicular to the bounding surface of the conductor.

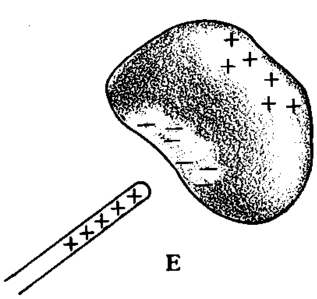


Figure 20-21

4. The induced charge per unit area σ on the conducting surface is related to the electric field E at the surface by

$$E = \frac{\sigma}{\epsilon_0} \tag{20-16}$$

Property 1 may be established as follows. Suppose that the electric field inside the conductor were not zero. Then the electrons there would experience an unbalanced force and accelerate in a direction opposite to this electric field. This would imply, contrary to our hypothesis, that the situation is not static. Thus it follows that under conditions of electrostatic equilibrium—our only concern for the moment—the electric field must vanish everywhere in the interior of a conductor.

We may understand this physically in the following way. When the conductor is first introduced into the electric field, the field lines penetrate the conductor and exert forces on the mobile electrons in its interior. In response to this force, the electrons tend to travel in a direction opposite to the field and thus accumulate on a surface of the conductor. As more electrons collect at this surface, they repel other electrons, thereby counteracting the effect of the original field. Ultimately, these surface electrons produce an electric field which entirely cancels the original field. By way of example, in Figure 20-22 we present the initial and final situations when a rectangular slab of metal is introduced into a uniform electric field **E**. Figure 20-22a shows the initial nonequilibrium situation in which the field lines penetrate the conductor so that the electrons experience a force acting to the left. Finally, after a sufficient amount of negative charge has accumulated on the left face of the conductor (thereby leaving a positive charge on the opposite face) the electric field in the interior becomes zero. Only then does this migration of electrons stop. The time involved in going from the initial to the final configuration of zero electric field is typically of the order of 10^{-14} second. This time is so short that for practical purposes this charge rearrangement can be thought of as taking place instantaneously.

To establish property 2, that the interior of a conductor is charge free, consider in Figure 20-23 an arbitrary closed Gaussian surface S contained

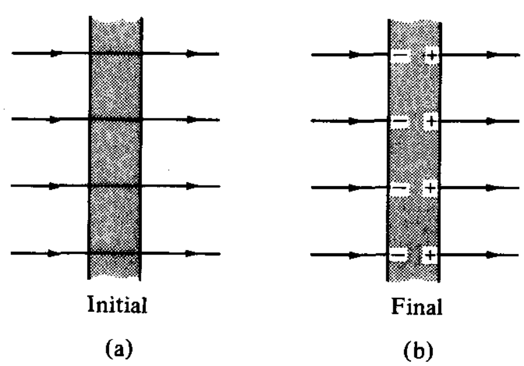


Figure 20-22

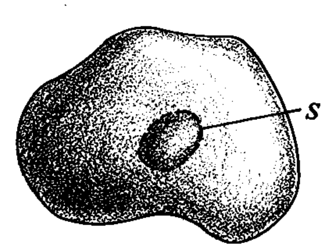


Figure 20-23

wholly within the interior of a conductor. According to property 1, the electric field vanishes everywhere on S and hence so must the flux Φ_S out of S. Applying Gauss' law to S, we conclude that the total charge inside S is also zero. Property 2 now follows from the fact that S is arbitrary. Note that this argument does *not* rule out the possibility of a surface charge on the conductor.

To establish property 3, that the electric field must be directed perpendicularly outward at each point of the conducting surface, we argue as follows. Suppose that at some point on the conducting surface the electric field had a component along the surface. Then the electrons near this point would experience a force and tend to move along the surface of the conductor. This contradicts the hypothesis that static conditions prevail. It follows then that the electric field at the surface of a conductor must be perpendicular to the surface.

Finally, let us establish property 4, according to which the electric field and the surface charge on a conductor are proportional to each other. To this end, consider, in Figure 20-24, a portion of the surface of a conductor in an electrostatic field **E** and suppose that the conductor has a surface charge σ.

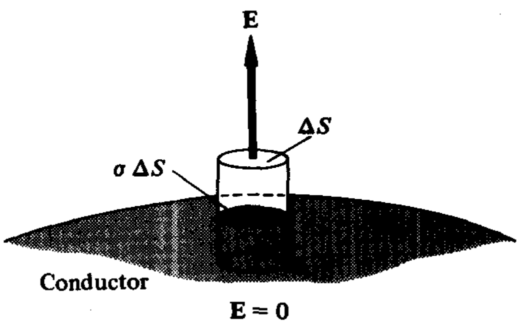


Figure 20-24

Let us construct a Gaussian surface in the form of a very small cylinder of cross-sectional area ΔS and with axis perpendicular to the surface at the given point. The total charge inside this Gaussian surface is $\sigma \Delta S$. Because of the facts that **E** is perpendicular to the surface of the conductor and vanishes inside, it follows by use of (20-12) that the total flux out of the end faces of the cylinder is simply $E \Delta S$. The flux out of the sides vanishes since **E** and Δ**S** are perpendicular along here. Applying Gauss' law we obtain

$$E \,\Delta S = \frac{1}{\epsilon_0} \sigma \,\Delta S$$

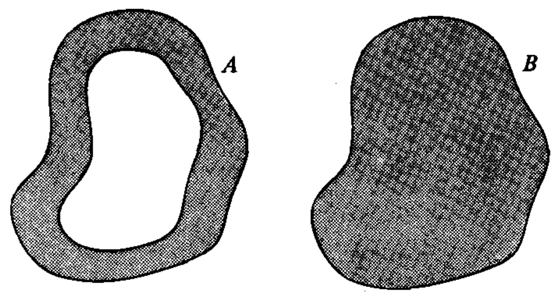
and this simplifies to

$$E = \frac{1}{\epsilon_0} \sigma$$

The validity of property 4 is thereby established.

Example 20-10 Show that the electric field in the interior of a hollow, charge-free, and closed conductor vanishes.

Solution To establish that the interior of the hollow, charge-free conductor A, in Figure 20-25a, has no electric field, consider first in Figure 20-25b a second conductor B, which is identical to A in all respects except that it is solid. According to the above arguments, regardless of the external charge, the electric field and the charge density vanish everywhere inside B. Thus, we may cut out any part of the interior of B without affecting either the external field lines or the surface charge. Since in this process of removing an interior portion of B we can produce the hollow conductor A, the desired result is established.



(a) (b)

Figure 20-25

This feature that the electric field in the interior of a hollow conductor vanishes is known as the principle of *electrostatic shielding*. If for any reason it is desired to shield out from a given region of space any stray electric fields that might exist, we may simply surround the region with a conducting shell and in this way obtain the desired shielding.

20-9 Experimental confirmation of the inverse-square law

Suppose a charge q is placed on a solid conductor. According to the arguments of Section 20-8, this charge must distribute itself on the outer surface of the conductor in such a way that the electric field inside vanishes. For if it did not, electrons would continue to migrate, and this is impossible under conditions of electrostatic equilibrium. Moreover, since $\mathbf{E}$ vanishes inside, it follows from Gauss' law that there can be no net charge anywhere in the interior of the conductor either. Thus we conclude that any charge placed on a conductor can reside only on its outer surface.

In a similar way, if a charge q is placed on a hollow closed conductor, it will also distribute itself on its *outer* surface. The electric field in this case must vanish not only in the interior of the conductor but in the cavity enclosed by the conductor as well. Thus by use of arguments similar to those presented in Example 20-10, it is easy to see that there can be no charge whatsoever on the inner surface of the conductor either.

Let us emphasize that in these arguments which establish that all charge placed on a conductor must reside on its outer surface, Gauss' law plays a decisive role. Now, as noted previously, Gauss' law is valid only for an

inverse-square law, so if the exponent "2" in Coulomb's law did not have precisely this value, Gauss' law would be invalid. Hence if charge were placed on a hollow conductor, and if any of this charge were ever detected anywhere but on the outer surface, then a violation of Gauss' law would have been observed. In turn, this would imply that the exponent in Coulomb's law could not have the precise value of 2 that it does indeed have.

Even before the enunciation of Gauss' law in the nineteenth century, experiments designed to show that charge can reside only on the outer surface of a conductor were carried out by many people, including such notables as Benjamin Franklin, Henry Cavendish, Michael Faraday, and later on even Maxwell himself. The most accurate measurement in the nineteenth century, of the value for this exponent, was made by Maxwell, who established to an accuracy of 50 parts per million that indeed it had the value 2. That is, Maxwell showed that to the accuracy of his experiments the value for the exponent must fall in the range 1.99995–2.00005. More recently, Plimpton and Lawton (1936) improved the accuracy of these experiments even more and concluded that to 2 parts per billion, the exponent in Coulomb's law has the precise value of 2. At the present time, therefore, this exponent is bounded by the values 1.999999998 and 2.000000002, and the precise value of 2 is very plausible.

The experimental techniques underlying precision measurements of the exponent in Coulomb's law are all basically very similar and relatively easy to understand qualitatively. Figure 20-26a shows a hollow, spherical conduc-

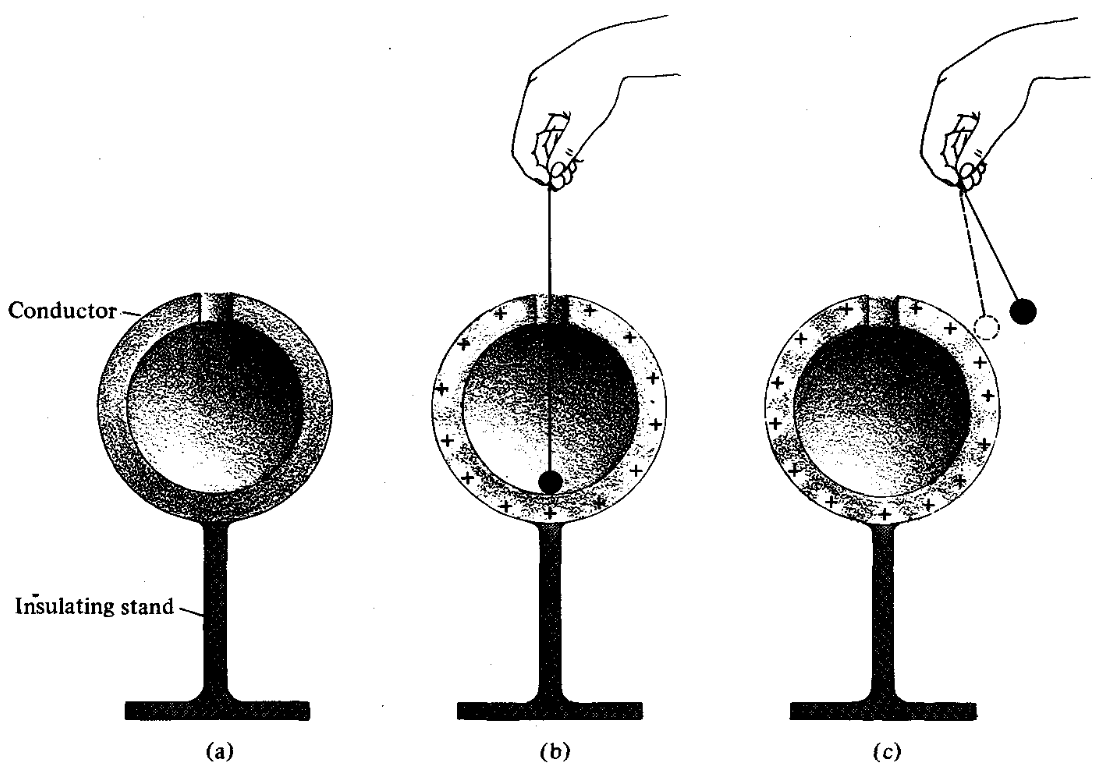
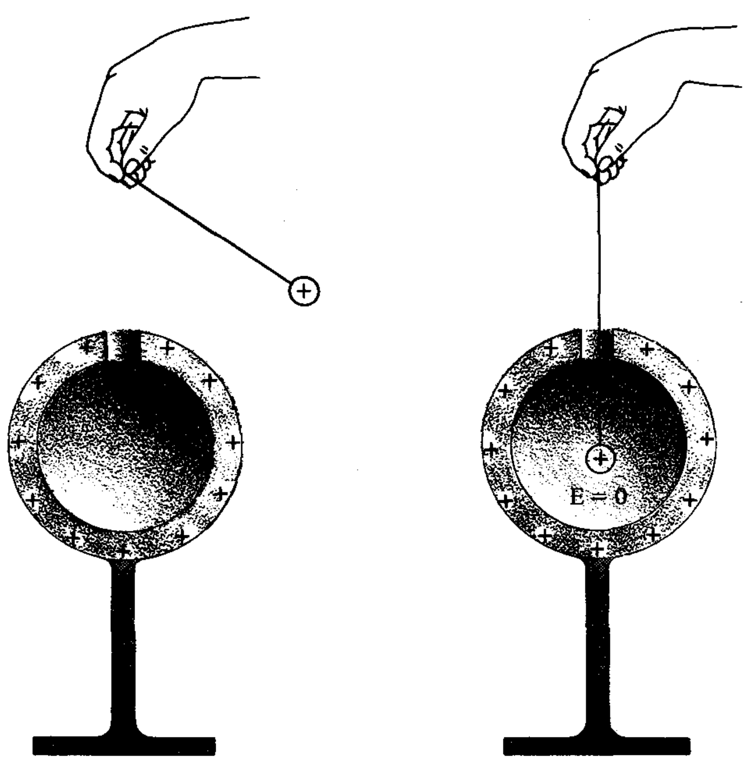


Figure 20-26

tor mounted on an insulating stand and with a small opening at the top so that small objects can be placed in the hollow part of the sphere. Suppose, as in Figure 20-26b, that we place a charge on the spherical conductor, lower into it a small, electrically neutral, conducting sphere, and let it come into contact with the inner surface of the hollow conductor. No measurable amount of charge ever accumulates on the small sphere. This shows that there is no charge on the inner surface of the hollow conductor. By contrast, if, as in Figure 20-26c, the small sphere is touched to the outer surface of the hollow conductor, it will immediately acquire an electric charge and will be repelled away. This confirms that all excess charge placed on a conductor resides entirely on its outer surface.

In a similar way we can establish that there is no electric field in the cavity. If, as before, a charge is placed on the hollow sphere, and a particle of very small charge is brought near its outer surface, then as shown in Figure 20-27a it will experience an electric force. This shows that outside the conducting shell there exists an electric field. By contrast, if this same charged conducting body is lowered into the cavity, it experiences no electric force, thus demonstrating that the electric field in the cavity is zero. By refinements of experiments of this sort the validity of Gauss' law and the fact that the exponent in Coulomb's must have the value of 2 has been confirmed to a very high degree of precision.



(a) (b)

Figure 20-27

20-10 Applications of Gauss' law

The electric fields associated with certain charge distributions having a high degree of symmetry may sometimes be obtained by use of Gauss' law. In this section we illustrate the method—and incidentally derive a number of useful formulas—by reference to three examples. The procedure used in each case involves the construction of a Gaussian surface of a nature such that the electric flux out can be expressed algebraically in terms of the electric field at the surface. An application of Gauss' law then yields an explicit formula for the electric field.

Example 20-11 Calculate the electric field due to an infinite nonconducting plane which has a uniform charge per unit area σ.

Solution Consider, in Figure 20-28, a Gaussian surface in the form of a very small cylinder of cross-sectional area ΔS and with its axis perpendicular to the plane. Assume the plane bisects the cylinder. The total charge inside the cylinder is $\sigma\, \Delta S$.

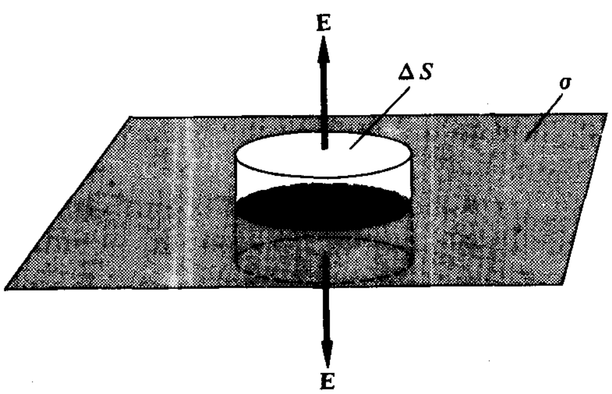


Figure 20-28

To calculate the electric flux, note first that, by symmetry, the electric field must be perpendicular to the plane. For if at some point the field had a component along the plane, this would imply more charge to one side of this point than the other and this is not possible for an infinite plane. Hence there are no contributions to the electric flux from the sides of the cylinder in the figure. On the other hand, there is a contribution to this flux at both the upper and lower faces. Again, by symmetry, the electric field **E** is the same on both sides of the plane, so the electric flux out of each of the end faces of the cylinder is $E\Delta S$. Hence the total flux out of the cylinder is $2E\, \Delta S$ and, by Gauss' law,

$$2E\, \Delta S = \frac{\sigma\, \Delta S}{\epsilon_0}$$

Solving for E we obtain

$$E = \frac{\sigma}{2\epsilon_0} \tag{20-17}$$

which is identical to (20-9), as it must be.

Example 20-12 A very long nonconducting cylinder of radius a contains a uniform charge per unit volume ρ_0. Calculate the electric field at a point outside the cylinder.

Solution Let us construct a Gaussian surface in the form of a cylinder of radius r ($>a$) of length l and coaxial with the charged cylinder (the dashed surface in Figure 20-29).

Again, because of the symmetry, the electric field **E** must be directed along the radial direction and can vary only with the distance r from the axis of the cylinder. Since for a radial field of this type the electric flux out of the end faces vanishes, it follows that the total flux Φ out of the Gaussian surface is

$$\Phi = \oint \mathbf{E} \cdot d\mathbf{S} = E \int dS = 2\pi r l E$$

The total charge inside is $\rho_0 \pi a^2 l$ and thus by Gauss' law

$$2\pi r l E = \frac{1}{\epsilon_0} \pi a^2 l \rho_0$$

Solving for E we obtain

$$E = \frac{a^2 \rho_0}{2\epsilon_0} \frac{1}{r} \qquad (r \geq a) \tag{20-18}$$

If we think of the charged cylinder as a line charge, then its charge per unit length λ is $\pi a^2 \rho_0$. Making this substitution in (20-18), we regain (20-6).

In the problems it is established that the field *inside* the charged cylinder in Figure 20-29 is

$$E = \frac{r\rho_0}{2\epsilon_0} \qquad (r \leq a) \tag{20-19}$$

Example 20-13 Show that the electric field outside of a uniformly charged sphere is the same as if the total charge of the sphere were concentrated at its center.

Solution Consider, in Figure 20-30, a sphere of uniform charge of radius a and thus of total charge $Q_0 = (4\pi/3)a^3\rho_0$, with ρ_0 the charge density. By symmetry, the electric field must be radial and can vary only with the radial coordinate r. Hence the flux out of the spherical Gaussian surface of radius $r(>a)$ is $4\pi r^2 E$, where E is the electric field at a distance r from the sphere. By Gauss' law we have

$$4\pi r^2 E = \frac{1}{\epsilon_0} Q_0$$

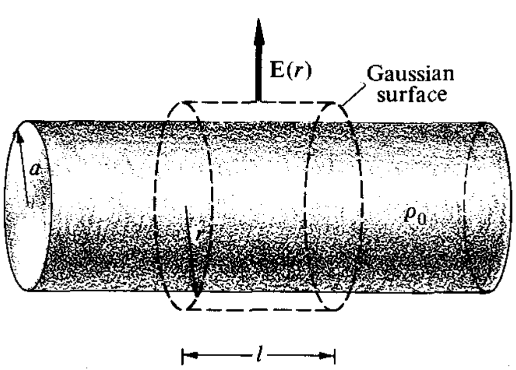


Figure 20-29

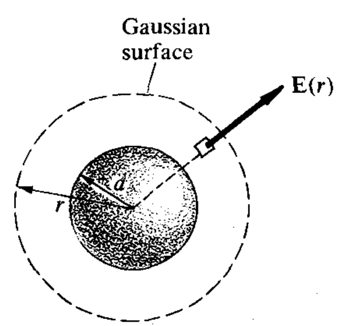


Figure 20-30

and thus we obtain the sought-for result

$$E = \frac{Q_0}{4\pi\epsilon_0}\frac{1}{r^2} \qquad (r \geq a) \tag{20-20}$$

If we take a spherical Gaussian surface with radius $r(<a)$, it is shown in the problems that inside the sphere the electric field is

$$E = \frac{\rho_0}{3\epsilon_0} r \qquad (r \leq a) \tag{20-21}$$

These results are plotted in Figure 20-31.

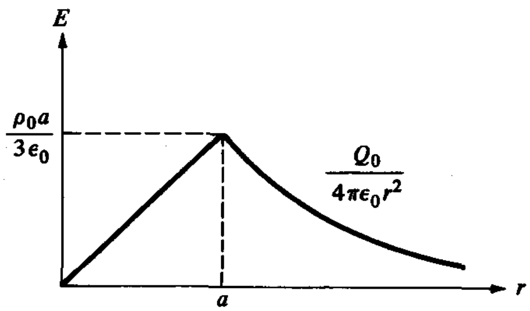


Figure 20-31

20-11 Summary of important formulas

The electric field **E** due to a distribution of charged particles is defined by

$$\mathbf{E} = \lim_{q_0 \to 0} \frac{1}{q_0}\mathbf{F} \tag{20-2}$$

where **F** is the force on a test particle of charge q_0 placed at the given field point. For a single particle of charge q, the electric field is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{\mathbf{r}} \tag{20-3}$$

where r is the distance of the particle from the field point and $\hat{\mathbf{r}}$ is a unit vector directed from the particle to the field point. For a collection of N particles of charges $q_1, q_2, \ldots, q_N$, **E** is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{q_i}{r_i^2}\hat{\mathbf{r}}_i \tag{20-5}$$

where r_i is the distance between the ith particle and the field point and $\hat{\mathbf{r}}_i$ is a unit vector directed along the line joining the ith particle to the field point.

The electric flux $\Delta\Phi$ through a small vectorial area element $\Delta\mathbf{S}$ is defined by

$$\Delta\Phi = \mathbf{E} \cdot \Delta\mathbf{S} = E\,\Delta S \cos\theta \tag{20-10}$$

where θ is the angle between the normal to the area element and the direction

of the electric field. For a finite surface S, the flux Φ_S is

$$\Phi_S = \sum \mathbf{E} \cdot \Delta\mathbf{S} \tag{20-12}$$

where the sum is to be carried out over every element $\Delta\mathbf{S}$ of the surface. In the limit as these area elements tend to zero, this sum approaches the surface integral

$$\Phi_S = \int_S \mathbf{E} \cdot d\mathbf{S} \tag{20-13}$$

Gauss' law states that

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} Q \tag{20-15}$$

where Q is the total charge enclosed by the closed surface S.

QUESTIONS

1. Why is it useful to introduce the notion of the electric field?
2. Explain why the definitions in (20-1) and (20-2) are equivalent if no conductors or insulators are present.
3. Consider an electric field line that lies entirely in the x-y plane. Why is the slope dy/dx at each point of this line given by

$$\frac{dy}{dx} = \frac{E_y}{E_x}$$

where E_y and E_x are the components of the electric field at the given point? How could you get the equations for the field lines by use of this property?
4. Consider the electric field $\mathbf{E}$ due to a positively charged point particle. Is the magnitude of the force on a nearby conducting sphere which carries a charge q_0 greater than or smaller than $|q_0|E$? Consider both cases of positive and negative q_0.
5. In what way, if any, is your answer to Question 4 modified if the field $\mathbf{E}$ is due to a uniformly charged, nonconducting sphere?
6. Consider an experiment for measuring the electric field $\mathbf{E}$ in the region surrounding a charged, conducting sphere by probing with a particle of charge q_0. Explain why you would obtain different answers depending on whether the definition in (20-1) or (20-2) is used. What is the basic advantage of using the latter definition?
7. Let $\mathbf{E}$ represent the electric field associated with an isolated charged conducting sphere. Is the magnitude of the force on a nearby particle of charge q_0 greater or less than $|q_0E|$? Consider all possibilities for the signs of the two charges.
8. Using the fact that the electric field outside of a uniformly charged, nonconducting sphere is the same as if all of the charge were concentrated at its center, explain why the force on such a sphere when placed into an external field is also the same as if all of the sphere's charge is concentrated at its center. (*Hint*: Use Newton's third law.)
9. Explain in physical terms why the strength of the electric field due to an infinite plane of charge is independent of distance from the plane.
10. Show by use of symmetry arguments that the electric field at the upper face of the cylinder in Figure 20-28 is

equal and opposite to that on the lower face.

11. A charged particle is originally at rest in a uniform electric field. Explain why it will follow a field line in its subsequent motion. Will it follow a field line if it originally had a component of velocity at right angles to a field line?
12. Repeat Question 11, but suppose that this time the electric field lines are curved.
13. Consider the electric field at a given point on the surface of a uniformly charged sphere. Show by symmetry arguments that every other point on the surface of the sphere has the same value for the electric field. (*Hint*: Consider an arbitrary rotation of the sphere about its center and use the fact that the physical situation is not altered in this process.)
14. Repeat Question 13, but this time consider the field at those points that are at a distance r from the center of the sphere.
15. Repeat Question 13, but this time assume that the charge distribution in the sphere is not uniform but is still spherically symmetric.
16. Consider the field produced by a particle of charge q. What is the flux out of each of the following surfaces: **(a)** A sphere surrounding the charge? **(b)** A cube surrounding the charge? **(c)** The surface bounded by two spheres of radii r_1 and r_2 $(r_1 < r_2)$ concentric with the particle?
17. Consider the field produced by a dipole. What is the flux out of a cubical surface that completely surrounds the dipole? Does the normal component of the electric field vanish on the surface of the cube?
18. Explain why it is *not* possible to calculate the field of a dipole by use of Gauss' law alone.
19. A sphere contains a charge distribution that is everywhere positive and spherically symmetric. Show by use of Gauss' law that the field must be a maximum at the surface of the sphere. Compare with Figure 20-31.
20. A spherical hole is cut out of the center of a sphere that contains a uniform charge distribution. Why must the electric field in the cavity vanish? (*Hint*: Use Gauss' law and symmetry arguments.)
21. A particle of charge q is placed in the interior of a hollow conductor, which is grounded. What is the *total* charge on the inner surface of the conductor? What is the charge on the outer surface of the conductor? Is there an electric field outside?
22. Why is it necessary that the charge on the particle in Figure 20-27b be vanishingly small? Why must there be a force on it if the charge is not small? Would this imply that the electric field in the cavity surrounded by a charged conducting shell is not zero? Explain.
23. Present an argument to show that Gauss' law implies that electric field lines must be continuous.
24. A particle of charge q is inside the cavity of a grounded conducting shell. Explain why the total charge induced on the conductor is $-q$. (*Hint*: What is $\mathbf{E}$ inside the conductor?)

PROBLEMS

1. Calculate the strength of the electric field produced by a particle of charge 5.0 μC at distances of 10^{-2}, 10^{-1}, and 10 meters from the particle.
2. What is the force, magnitude and direction, exerted on a particle of charge $-2.0\ \mu$C when it is in a uniform field of strength 10^3 N/C?
3. Two particles, each of charge

+3.0 μC are located at the points $(2, 0, 0)$ and $(-2, 0, 0)$ in a certain coordinate system. Calculate the magnitude and direction of the field at the following points, assuming that all distances are measured in meters: **(a)** $(0, 0, 0)$; **(b)** $(1, 0, 0)$; and **(c)** $(3, 0, 0)$.

4. Repeat Problem 3, but this time calculate the field at the points **(a)** (0, 1, 0); **(b)** (0, −1, 0); and **(c)** (2, 2, 0).

5. A dipole of dipole moment **p** consists of two particles of charges $+q$ and $-q$ and located in a certain coordinate system at the respective points $(a, 0, 0)$ and $(-a, 0, 0)$. In terms of these parameters the dipole moment **p** is defined by

$$\mathbf{p} = 2aq\mathbf{i}$$

where **i** is a unit vector along the positive x-axis. Calculate the electric field **E** of this dipole at the following points: **(a)** $(2a, 0, 0)$ and **(b)** $(-2a, 0, 0)$.

6. Show by use of the results of Problem 5 and Example 20-2 that the torque $\boldsymbol{\tau}$ on a dipole of moment **p** in a uniform field **E** may be expressed in the form

$$\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E}$$

***7.** Use the result of Example 20-2 to show that:

(a) The dipole is in stable equilibrium when $\theta = 0$, that is, when it is lined up with the field.

(b) The dipole is in unstable equilibrium when $\theta = \pi$. (*Note*: Stability occurs, if the forces and torques are of such a nature that the system tends to return to its original equilibrium state whenever it is slightly perturbed. Otherwise, it is unstable.)

8. As measured in a certain coordinate system, $2N$ particles, each of charge q, are located at the points with coordinates $\{(na, 0, 0)\}$, where $n = \pm 1, \pm 2, \ldots, \pm N$. Show that the electric field **E** at the point $(0, d, 0)$ is

$$\mathbf{E} = \mathbf{j}\frac{qd}{2\pi\epsilon_0}\sum_{n=1}^{N}\frac{1}{[d^2 + n^2a^2]^{3/2}}$$

9. What charge per unit area σ must be placed on a very large plane so that the associated electric field has a strength of 100 N/C?

10. A particle of charge q and mass m is suspended from an insulating thread and hangs near a vertical infinite plane, which has a charge per unit area σ (Figure 20-32).

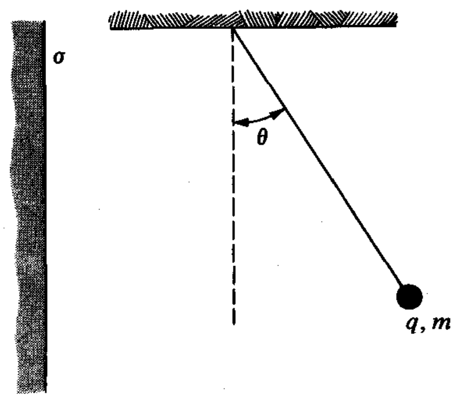


Figure 20-32

(a) Show that the angle θ which the thread makes with the vertical is given by

$$\tan\theta = \frac{q\sigma}{2\epsilon_o mg}$$

(b) Calculate the tension T in the string.

11. Consider a proton in a certain uniform electric field **E**. **(a)** What must be the strength of this electric field so that it just balances the earth's gravitational field? **(b)** At what distance from a particle of charge 10 μC is a field of this strength achieved?

12. A straight line charge of length l carries a uniform charge per unit

length λ. Show that the magnitude of the electric field at a point on the axis of the line charge and at a distance x from the nearer end is

$$E = \frac{\lambda l}{4\pi\epsilon_0} \frac{1}{x(l+x)}$$

13. A line charge of length $2l$ consists of two parts: one half carrying a charge per unit length $+\lambda$ and the other half the charge per unit length $-\lambda$. Calculate the electric field at a distance y from the line charge and along its perpendicular bisector.
14. Calculate the magnitude and direction of the electric field at a perpendicular distance of 50 cm from a very long wire that carries a uniform charge of 2.0 μC/m.
15. An electron is at a distance of 2.0 cm from a very long wire and approaching it with an acceleration of 1.5×10^{13} m/s^2. What is the charge per unit length on the wire?

*16. A circular hole of radius a is cut out of an infinite plane that carries a charge per unit area σ. Calculate the electric field at a point along the axis of the hole and at a distance b from the plane. (*Hint*: Use Example 20-6 and the principle of superposition, according to which this field is the sum of the fields due to an infinite plane of charge $+\sigma$ plus that due to a disk of radius a carrying a charge $-\sigma$.)

17. A line charge of charge per unit length λ is in the form of a semicircle of radius a. Calculate the electric field at the center of the circle.
18. Calculate the electric field for the line charge in Problem 17, but this time take the field point at a distance b vertically above the center of the circle.
19. A particle of charge $-q_0\,(q_0 > 0)$ and mass m is constrained to motion along the axis of a circular line charge of radius a and of charge per unit length λ (>0).
    (a) What is the force acting on the particle when it is at the center of the ring?
    (b) Show that if it is displaced from this position (along the axis of the ring), then it executes simple harmonic motion with period
    $$T = 2\pi\left(\frac{2\epsilon_0 m a^2}{q_0\lambda}\right)^{1/2}$$
20. A disk of radius a carries a surface charge per unit area σ, which varies with radius r as
    $$\sigma = \frac{\sigma_0}{a} r$$
    where σ_0 is a positive constant.
    (a) What is the total charge on the disk?
    (b) Calculate the electric field at a distance b from the plane of the disk and along its axis. You may use the indefinite integral
    $$\int \frac{x^2\,dx}{(x^2+b^2)^{3/2}} = -\frac{x}{(x^2+b^2)^{1/2}} + \ln[x + (x^2+b^2)^{1/2}]$$
21. A disk of radius R_0 carries a charge per unit area σ and has a hole of radius a cut out of its center. By an appropriate modification of (20-8) calculate the electric field at a point on the axis of the disk and at a distance b away from its center.
22. Show by use of (20-7) that the field on the axis of a uniformly charged circular loop of radius a has the magnitude
    $$E = \frac{Qb}{4\pi\epsilon_0[a^2+b^2]^{3/2}}$$
    where Q is the total charge on the loop and b is the distance from the center of the loop. Does this reduce to the expected formula in the limit $b \gg a$?

*23. A cylinder of length l and radius R has a uniform charge per unit volume ρ_0. Show that the electric field at a point on the axis and at a distance b from the nearer end face is directed along the axis and has the magnitude

$$E = \frac{\rho_0}{2\epsilon_0}\left\{l + (R^2 + b^2)^{1/2} - [R^2 + (b+l)^2]^{1/2}\right\}$$

(*Hint*: Find the field $d\mathbf{E}$ due to a thin disk and integrate.)

24. Calculate the net charge inside a closed surface if the flux out of the surface is 5.0×10^4 N-m²/C.

*25. A particle of charge q is at the point $(-a, 0, 0)$ in a certain coordinate system (Figure 20-33).

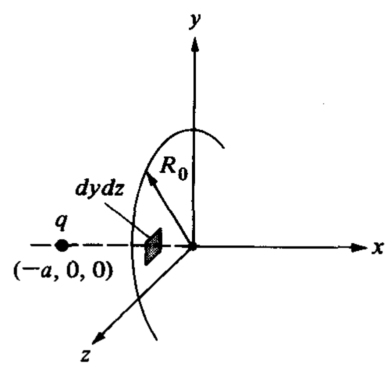


Figure 20-33

(a) Show that the flux $d\Phi$ through the element of area $dy\,dz$ at the point $(0, y, z)$ in the y-z plane is

$$d\Phi = \frac{qa}{4\pi\epsilon_0}\frac{1}{(a^2 + y^2 + z^2)^{3/2}}\,dy\,dz$$

(b) Calculate the total flux through a circle of radius R_0, centered at the origin and lying in the y-z plane. *Hint*: Transform to polar coordinates and evaluate

$$\Phi = \frac{qa}{4\pi\epsilon_0}\int_0^{R_0} r\,dr \int_0^{2\pi} d\theta \frac{1}{(a^2 + r^2)^{3/2}}$$

(c) What is the total flux through the entire y-z plane?

26. Two spheres of radii R_1 and R_2 ($< R_1$) are concentric with a particle of charge q.
   (a) What is the flux through the smaller sphere?
   (b) What is the flux through the larger sphere?
   (c) By use of your results to (a) and (b) calculate the flux through the closed surface consisting of these two spheres. Is your answer consistent with Gauss' law?

27. A small body carrying a charge of 2.5 μC is suspended in a cavity inside an uncharged conducting body.
   (a) What is the electric field inside the conducting medium?
   (b) By constructing an appropriate Gaussian surface, calculate the total charge induced on the inner walls of the cavity.
   (c) What is the total charge on the outer surface of the conductor?

28. Consider an infinite cylinder of radius a, which carries a uniform charge per unit volume ρ_0. Construct a Gaussian surface as in Figure 20-29, but this time with $r < a$.
   (a) Why must the electric field $\mathbf{E}$ be everywhere perpendicular to the curved part of the Gaussian surface?
   (b) What is the total charge inside the Gaussian surface?
   (c) Calculate the flux through the Gaussian surface and thus derive (20-19).

29. An infinitely long cylinder of radius a carries a uniform charge per unit volume $-\rho_0$ ($\rho_0 > 0$) and is surrounded by a grounded conducting shell of radius b that is coaxial with the cylinder; see Figure 20-34. Calculate, by use of suitable Gaussian surfaces: (**a**) the electric field for $r \leq a$; (**b**) the electric field for $a \leq r \leq b$; and (**c**) the electric field for $r \geq b$.

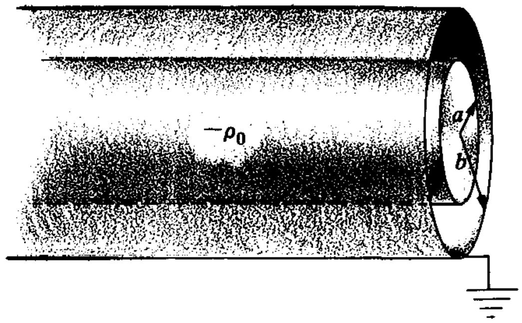


Figure 20-34

30. For the physical system of Problem 29, calculate the charge per unit area on the surface of the grounded conducting cylinder.

***31.** An infinitely long cylinder of radius a has a charge per unit volume

$$\rho = \rho_0\left(\frac{1}{2} - \frac{r^2}{a^2}\right)$$

where ρ_0 is a positive constant and r is the radial distance. To calculate the electric field for $r \leq a$, construct a Gaussian surface in the form of a cylinder of radius r and length l.

(a) Show that the total charge inside the Gaussian surface is

$$Q(r) = \frac{\pi l \rho_0}{2a^2} r^2(a^2 - r^2)$$

(b) Calculate the electric field inside the cylinder.

32. Calculate the field outside of the cylinder of Problem 31 by constructing an appropriate Gaussian surface and using Gauss' law.

33. Derive (20-21), which gives the electric field at a point inside of a uniformly charged sphere of radius a. (*Hint*: Apply Gauss' law to a spherical surface of radius r ($< a$).)

34. By use of (20-20) and (20-21) prove that the electric field at a distance r from the center of a uniformly charged sphere is due to all charge no farther than r from the center of the sphere.

35. A nucleus of mass number A and containing Z protons may be thought of as a uniformly charged sphere of radius R_0 given (in meters) by

$$R_0 = 1.1 \times 10^{-15} A^{1/3}$$

(a) In terms of A and Z and the quantum of charge e, calculate the electric field outside of this nucleus, at a distance r from its center.

(b) Calculate the electric field inside the nucleus.

(c) Make a plot of the electric field as a function of r for the special nucleus $^{235}_{92}U$.

***36.** A very small cylindrical hole is made along a diameter through a uniformly charged sphere of density ρ_0 (>0) of radius a. Assume that the electric field inside this hole is the same as it would be if the hole were not present. Suppose now that a particle of charge $-q_0$ ($q_0 > 0$) and mass m is dropped into the hole. Assume no friction. (a) Show by use of (20-21) that the particle will move back and forth with simple harmonic motion. (b) Calculate the period of this motion.

***37.** A sphere of radius a contains a uniform charge density ρ_0 and is surrounded by a concentric spherical shell of inner radius b ($> a$) and outer radius c and containing a uniform charge density $-\rho_0$. See Figure 20-35. By constructing suitable Gaussian surfaces calculate the electric field for values of r satisfying:

(a) $r \leq a$;

(b) $a \leq r \leq b$;

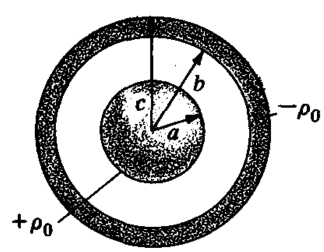


Figure 20-35

(c) $b \leqslant r \leqslant c$;
(d) $r \geqslant c$.

*38. Consider a spherically symmetric distribution of charge characterized by the charge density $\rho(r)$.
(a) Show by arguments of symmetry that the electric field must be purely radial and that it can depend only on r.
(b) Show that the electric field at the radial distance r is

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q(r)}{r^2}$$

where $Q(r)$ is the total charge contained within a sphere of radius r:

$$Q(r) = 4\pi \int_0^r r^2\, dr\, \rho(r)$$

39. Apply the result of Problem 38 to show that the electric field vanishes in a spherical cavity cut out of the center of a uniform distribution of charge.

40. A particle of charge q is placed at the center of a spherical conducting shell of inner radius a and outer radius b. Calculate the electric field for the following distances:
(a) $r < a$;
(b) $a \leqslant r \leqslant b$;
(c) $r > b$.

41. Consider the same situation as in Problem 40. (a) What is the surface charge density σ on the inner and outer surfaces of the conducting sphere? (b) Repeat (a), but this time assume that the sphere is grounded.

42. Consider the same situation as in Problem 40, but suppose that, in addition, a charge Q is placed on the conductor. (a) What is the total charge on the inner surface of the conducting sphere? (b) What is the total charge on the outer surface?

43. Calculate the electric field (in terms of q and Q) everywhere for the situation described in Problem 42.

*44. Consider, in Figure 20-36, a sphere of radius a containing a uniform charge density ρ_0 out of which is cut a spherical hole of radius c whose center is located at the vectorial distance **b** from the center of the big sphere.

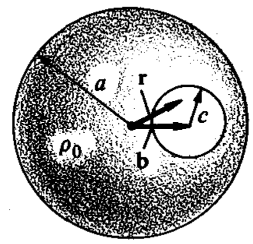


Figure 20-36

(a) Why can we express the total field **E** in the form

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$$

where $\mathbf{E}_1$ is the field due to a uniformly charged sphere of density ρ_0 and radius a, and $\mathbf{E}_2$ is the field produced by a uniformly charged sphere of radius c located at **b** and carrying a charge density $-\rho_0$?
(b) Show that if **r** is any point in the cavity, then

$$\mathbf{E}_1 = \frac{\rho_0}{3\epsilon_0}\mathbf{r}$$

(c) Show that $\mathbf{E}_2$ is

$$\mathbf{E}_2 = \frac{-\rho_0}{3\epsilon_0}(\mathbf{r} - \mathbf{b})$$

if **r** is any point in the cavity.
(d) By combining these results, show that the electric field **E** in the cavity is uniform and has the value

$$\mathbf{E} = \frac{\rho_0}{3\epsilon_0}\mathbf{b}$$

21 The electrostatic potential

In every explanation of natural phenomena we are compelled to leave the sphere of sense perception and to pass to things which are not the objects of sense and are defined only by abstract conceptions.

H. L. F. VON HELMHOLTZ (1821-1894)

21-1 Introduction

The electrostatic field has two very striking and useful mathematical properties. One of these we studied in Chapter 20, and is that it satisfies Gauss' law. The other has to do with a way of characterizing it in terms of a certain scalar quantity called the *electrostatic potential.* This important quantity is the subject matter of this chapter.

Because of the fact that the electric field associates with each point in space a vector—that is, both a magnitude and a direction—it is known as a *vector field.* By contrast, the electrostatic potential is a *scalar field* by virtue of the fact that it associates with each point in space a scalar quantity. We might expect therefore that if a quantitative description of electrostatic phenomena in terms of such a potential function were possible it would be considerably simpler than that in terms of the electric field itself. This is indeed the case, as we shall confirm in this chapter.

21-2 Work and the electric field

Consider, in Figure 21-1, a region of space in which there exists an electric field $\mathbf{E}$ and suppose that an external agent moves a particle of charge q_0 very slowly along a curve C from a point A to a point B. To achieve this, he must exert on the particle a certain force $\mathbf{F}$, which is equal and opposite to the electric force $q_0\mathbf{E}$ on it at each point of its path.

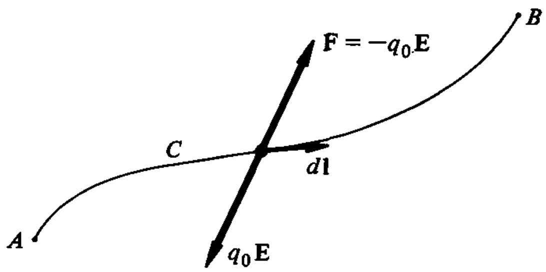


Figure 21-1

According to the definition in Chapter 7, the work W_{AB} that the agent carries out in moving the particle from A to B along a curve C is obtained by dividing the path into a sequence of infinitesimal displacements $\{\Delta\mathbf{l}\}$ and adding together the works $\mathbf{F}\cdot\Delta\mathbf{l}$ carried out along each of these. The work W_{AB}, which is the limit of this sum as the magnitudes of the displacements $|\Delta\mathbf{l}|$ tend to zero, is them defined to be the line integral

$$W_{AB} = \int_A^B \mathbf{F}\cdot d\mathbf{l} \tag{21-1}$$

where $d\mathbf{l}$ is an infinitesimal element of path along C, and the integral is along the curve C from A to B. For the situation in Figure 21-1, since $\mathbf{F} = -q_0\mathbf{E}$ it follows that the work W_{AB} carried out may be expressed as

$$W_{AB} = -q_0\int_A^B \mathbf{E}\cdot d\mathbf{l} \tag{21-2}$$

where the constant q_0 has been taken out from under the integral sign. Recall from the discussion in Chapter 7 that, in general, the value of the integral in (21-2) depends not only on the endpoints A and B of the path, but also on the particular curve C connecting them. In anticipation of a fact to be discussed in Section 21-3 that the integral in (21-2) is actually independent of the path, all reference to the particular path C has been omitted from this formula.

For the special case that the path C joining A and B is a straight line, the line integral in (21-2) reduces to an ordinary integral. The situation here is precisely the same as that considered in Chapter 7.

Example 21-1 A particle of charge q_0 is near an infinite nonconducting plane of charge per unit area σ. Calculate the work required to move it from A to B in Figure 21-2 along:

(a) The straight-line path of length a connecting A and B, assuming this line makes an angle α with the line $\overline{AD}$, which is perpendicular to the plane.

(b) The straight-line paths $\overline{AD}$ and $\overline{DB}$.

Solution

(a) In Example 20-6 it was established that the electric field associated with an infinite, nonconducting plane is directed perpendicular to the plane and has the value $\sigma/2\epsilon_0$. Substitution into (21-2) yields

$$W_{AB} = -q_0 \int_A^B \mathbf{E} \cdot d\mathbf{l} = -q_0 \int_A^B \frac{\sigma}{2\epsilon_0}\, dl \cos\alpha$$

$$= \frac{-q_0\sigma}{2\epsilon_0} \cos\alpha \int_A^B dl = -\frac{q_0\sigma a \cos\alpha}{2\epsilon_0}$$

since $\mathbf{E}$ is constant, as is the angle α between $\mathbf{E}$ and $d\mathbf{l}$.

(b) In this case note first that along $\overline{DB}$, $\mathbf{E}$ and $d\mathbf{l}$ are perpendicular, and hence no work is carried out along this portion of the path. Along $\overline{AD}$, $\mathbf{E}$ and $d\mathbf{l}$ are parallel, and thus

$$W_{AB} = -q_0 \int_A^D \mathbf{E} \cdot d\mathbf{l} = -q_0 \int_A^D \frac{\sigma}{2\epsilon_0}\, dl = -\frac{q_0\sigma}{2\epsilon_0} \int_A^D dl$$

$$= -\frac{q_0\sigma a \cos\alpha}{2\epsilon_0}$$

since the path length $\overline{AD}$ is $a \cos\alpha$. It is interesting to note that this formula for the work is precisely the same as that found in (a). This confirms the fact that W_{AB} is independent of the path, at least in this case.

21-3 The electrostatic potential

Consider, in Figure 21-3, a region of space in which there exists an electrostatic field $\mathbf{E}$, and let A and B represent two fixed points in this region. Imagine calculating, by use of (21-2), the work W_{AB} required to take a particle of charge q_0 from point A to point B along each of the three paths C_1, C_2, and C_3. In general, since the electric field will assume different values along these paths, we might expect that W_{AB} would also vary depending on which of the paths C_1, C_2, or C_3 is chosen. However, this is found *not* to be

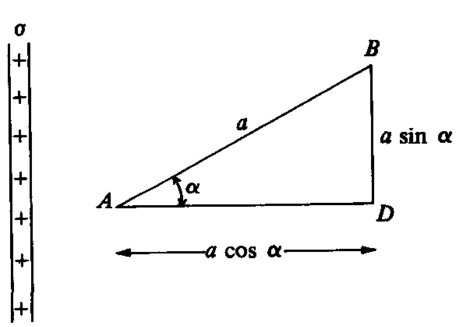


Figure 21-2

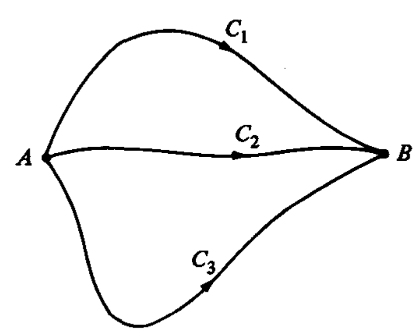


Figure 21-3

true. Instead, W_{AB} is found to be the *same* for *all* paths connecting points A and B. This property of the electrostatic field, that the integral in (21-2) is independent of the path, is described by saying that the electrostatic field is a *conservative field.* It is one of the two most important properties of this field; the other one is Gauss' law.

An equivalent way of describing the conservative nature of the electrostatic field is by the statement that no (zero) work is required to take a charged particle around any closed path. The equivalence of these two ways of characterizing a conservative field is established in the problems.

Now because of this fact that the integral

$$\int_A^B \mathbf{E} \cdot d\mathbf{l}$$

in (21-2) is independent of the path connecting the points A and B, it follows that the integrand $\mathbf{E} \cdot d\mathbf{l}$ must be a perfect differential. Hence, just as for the corresponding analysis of potential energy in Section 8-5, it follows that there must exist a scalar function of position V, called the *potential function,* with the property that its differential dV is related to the electric field $\mathbf{E}$ for an arbitrary infinitesimal displacement $d\mathbf{l}$ by

$$dV = -\mathbf{E} \cdot d\mathbf{l} \tag{21-3}$$

The minus sign here is conventional and has no other significance. Substituting this relation into the definition for the work W_{AB} in (21-2), we obtain

$$W_{AB} = -q_0 \int_A^B \mathbf{E} \cdot d\mathbf{l} = q_0 \int_A^B dV = q_0 V \Big|_A^B = q_0(V_B - V_A) \tag{21-4}$$

where the third equality follows since dV is a perfect differential, and where V_B and V_A are the values of the potential function at the points B and A, respectively. Note that W_{AB} is proportional to $(V_B - V_A)$ and thus has the desired property of depending *only* on the endpoints A and B and not on the path connecting them.

The minus sign, whose presence in (21-3) is only by convention, is important to keep in mind. Physically, it has the consequence that as a particle travels along the direction of a field line it goes to regions of lower potential. Thus the potential decreases as we recede from a positively charged particle or approach a negatively charged one.

An important consequence of the conservative nature of the electric field is that if for a given charge distribution the potential function is known everywhere, then to evaluate the work W_{AB} in (21-2) it is not necessary to evaluate the line integral directly. For, according to (21-4),

$$W_{AB} = q_0(V_B - V_A) \tag{21-5}$$

and since V is presumed known everywhere, W_{AB} can be obtained by simple substitution. This formula in (21-5) also suggests a physical interpretation for the *potential difference* $(V_B - V_A)$ as *the work per unit charge required to take a particle from A to B.*

The definition of the potential function in (21-3) is in terms of differentials and thus V is defined only up to an additive constant. This means in particular that only differences in potential, as in (21-5), for example, can be given a physical meaning. Following custom, we shall arbitrarily assign the value zero to the value for the potential function for the point at "infinity." With this convention, if point A in (21-5) is selected to be the point at infinity, so that $V_A = 0$, this relation may be interpreted by the statement that the *potential at a point is the work required to bring a unit charge from infinity to this point.*

The unit of potential, according to (21-5), is energy per unit charge, that is, the joule per coulomb. This unit potential is defined to be the *volt* (V), and thus, by definition,

$$1 \text{ V} = 1 \text{ J/C}$$

Thus it takes 1 joule of work to take a 1-coulomb particle through a potential difference of 1 volt. In terms of this unit, it follows from (21-3) that the volt per meter (V/m) is the unit of electric field strength. We shall express electric field strength in volts per meter from now on.

Note the distinction between the *italic* letter V for potential and the roman letter V for the unit of potential of the volt. As a general rule, the symbol for the unit V will usually be preceded by a numerical value.

Example 21-2 For the physical situation described in Example 21-1, calculate the potential difference between points:

(a) B and D.
(b) A and B.

Solution

(a) Along the path BD in Figure 21-2 the electric field is perpendicular to $d\mathbf{l}$ and hence $W_{BD} = 0$. It follows then from (21-5) that $V_B = V_D$ and thus $V_B - V_D = 0$.

(b) Comparing the result in Example 21-1 with (21-5), we find that

$$V_B - V_A = \frac{-\sigma a \cos \alpha}{2\epsilon_0}$$

21-4 Calculation of the electric field

A direct consequence of the defining relation in (21-3) is that if for a given charge distribution the potential function V is known, then the associated electric field $\mathbf{E}$ can be obtained by differentiation. The purpose of this section is to derive this important relation.

Suppose that for a given physical situation the potential function V is known everywhere, and that V_A represents its value at some given point A (Figure 21-4). If $\mathbf{E}$ is the electric field at this point, then, according to (21-3) the value of the potential at a point that is at an infinitesimal displacement $d\mathbf{l}$ from A is $(V_A + dV)$, with dV given by

$$dV = -\mathbf{E} \cdot d\mathbf{l} \tag{21-3}$$

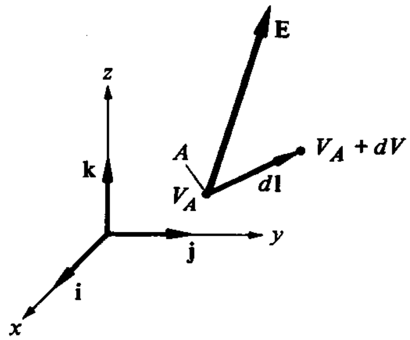


Figure 21-4

Note the important feature that for a fixed point A and a given value for $\mathbf{E}$ at that point, this formula gives the change in potential dV for *all* points at any infinitesimal displacement $d\mathbf{l}$ from A.

To derive the relation between $\mathbf{E}$ and V, let us assume a Cartesian coordinate system and make the particular choice $d\mathbf{l} = \mathbf{i}\,dx$ corresponding to a displacement along the x-axis. For this case the change in potential dV is

$$dV = -\mathbf{E} \cdot (\mathbf{i}\,dx) = -E_x dx \tag{21-6}$$

with E_x the component of $\mathbf{E}$ along the x-axis of the given system. Noting that for this change the variables y and z are kept fixed, we find on making use of the definition of a partial derivative that (21-6) is equivalent to

$$E_x = -\frac{\partial V}{\partial x}$$

Analogous relations may be obtained by considering displacements $d\mathbf{l} = \mathbf{j}\,dy$ and $d\mathbf{l} = \mathbf{k}\,dz$ along the y- and z-axes, respectively. Hence

$$E_x = -\frac{\partial V}{\partial x} \qquad E_y = -\frac{\partial V}{\partial y} \qquad E_z = -\frac{\partial V}{\partial z} \tag{21-7}$$

so that the *component of the electric field along any direction is the negative of the rate of change of the potential along that direction.*

From a practical point of view (21-7) is of considerable importance. By use of these formulas we can calculate the electric field for any charge distribution once its potential function V has been obtained. As will be seen below, the problem of calculating the potential function is, as a rule, appreciably simpler than that of calculating $\mathbf{E}$ directly.

Example 21-3 In Section 21-5 it is shown that the potential V associated with a dipole of moment p located at the origin of a certain coordinate system and oriented along the positive z-axis is

$$V = \frac{p}{4\pi\epsilon_0} \frac{z}{(x^2 + y^2 + z^2)^{3/2}}$$

Calculate the electric field everywhere for this dipole.

Solution To obtain **E** we need only evaluate the derivatives of V in accordance with (21-7). Thus

$$E_x = -\frac{\partial V}{\partial x} = -\frac{p}{4\pi\epsilon_0}\frac{\partial}{\partial x}\left[\frac{z}{(x^2+y^2+z^2)^{3/2}}\right]$$
$$= \frac{3p}{4\pi\epsilon_0}\frac{xz}{(x^2+y^2+z^2)^{5/2}}$$

and similarly for E_y

$$E_y = \frac{3p}{4\pi\epsilon_0}\frac{yz}{(x^2+y^2+z^2)^{5/2}}$$

For E_z it is necessary to apply the rule for differentiating a product. The result is

$$E_z = -\frac{p}{4\pi\epsilon_0}\frac{\partial}{\partial z}\frac{z}{(x^2+y^2+z^2)^{3/2}}$$
$$= -\frac{p}{4\pi\epsilon_0}\left[\frac{1}{(x^2+y^2+z^2)^{3/2}} - \frac{3z^2}{(x^2+y^2+z^2)^{5/2}}\right]$$
$$= -\frac{p}{4\pi\epsilon_0}\frac{1}{(x^2+y^2+z^2)^{5/2}}[x^2+y^2-2z^2]$$

Example 21-4 A solid conductor is placed into an electric field. Show that the potential associated with the resultant field must have the same value for every point inside and on the surface of the conductor.

Solution Let us first establish that every point at the surface of the conductor is at the same potential. To this end, consider in Figure 21-5 a small displacement $d\mathbf{l}_1$, which lies entirely *on* the surface of the conductor. If **E** is the electric field at this point, then the change dV in the potential associated with the displacement $d\mathbf{l}_1$ is given, according to (21-3), by

$$-\mathbf{E}\cdot d\mathbf{l}_1 = dV$$

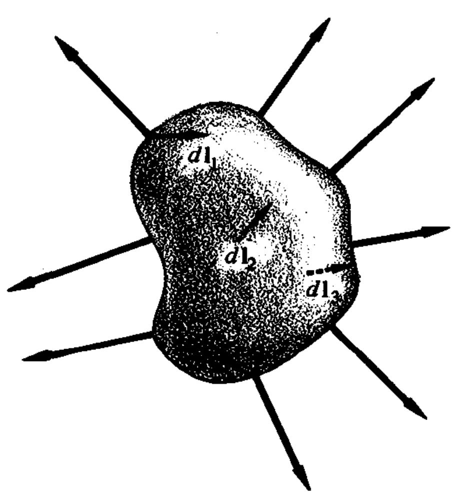


Figure 21-5

But the electric field on the surface of a conductor must be perpendicular to its surface. Hence, since $d\mathbf{l}_1$ lies on the surface, it follows that $\mathbf{E} \cdot d\mathbf{l}_1 = 0$, and thus $dV = 0$. In other words, as we go from one point on the surface of a conductor to a neighboring point, the change in the potential is zero. Since by repeated use of this argument we can establish that the change in potential is zero between any two points on the conducting surface, it follows that all points on the conducting surface must be at the same potential.

In a similar way, since for each point in the interior of the conductor the electric field vanishes, it follows that $\mathbf{E} \cdot d\mathbf{l}_2 = 0$ for any displacement $d\mathbf{l}_2$ lying entirely in the interior of the conductor. Thus, again, each point inside the conductor must be at the same potential. Finally, by considering a small displacement $d\mathbf{l}_3$ from the interior of the conductor to its surface, we conclude that *all* elements of the conductor must be at the same potential.

21-5 The potential function for a system of charged particles

In order for (21-7) to be useful for calculating electric fields, it is necessary to have available formulas for the potential function of physically interesting systems. In this section we consider the problem of calculating the potential function for a collection of charged particles. The associated problem of continuous distributions will be considered in Section 21-6.

Let us consider first the case of a single particle of charge q, as in Figure 21-6. Since the potential associated with the point at infinity is assumed to vanish, it follows by integrating both sides of (21-3) along the radial path in the figure from infinity to point P that the potential V at a distance r from the particle is

$$V = -\int_{\infty}^{r} \mathbf{E} \cdot d\mathbf{l} \qquad \textbf{(21-8)}$$

Now the electric field $\mathbf{E}$ at a distance s from the particle is

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0} \frac{1}{s^2} \hat{\mathbf{r}}$$

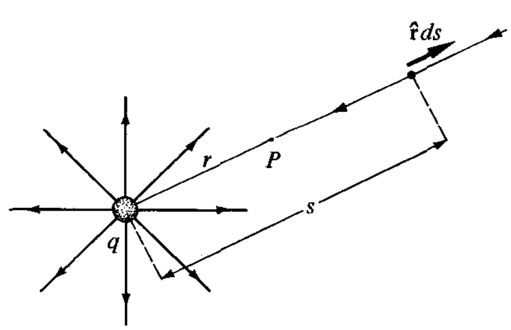


Figure 21-6

with $\hat{\mathbf{r}}$ a unit vector along the radial direction. An infinitesimal displacement $d\mathbf{l}$ of length ds along this direction is $\hat{\mathbf{r}}\,ds$. Substitution into (21-8) yields, for the potential at point P,

$$V = -\int_{\infty}^{r} \mathbf{E} \cdot d\mathbf{l} = -\frac{q}{4\pi\epsilon_0}\int_{\infty}^{r} \frac{ds}{s^2} = -\frac{q}{4\pi\epsilon_0}\left(-\frac{1}{s}\right)\Bigg|_{\infty}^{r}$$

where the last equality follows from the fact that $d(1/s)/ds = -1/s^2$. Thus we obtain the final formula

$$V = \frac{q}{4\pi\epsilon_0}\frac{1}{r} \qquad \textbf{(21-9)}$$

for the potential due to a charged particle.

The compatibility of this formula with (20-3) is easily established by use of (21-7) and the fact that $r = [x^2 + y^2 + z^2]^{1/2}$. The details are left as an exercise. Equivalently, since V in (21-9) varies only along the radial direction, it follows from (21-3) that $\mathbf{E}$ must also. And its magnitude E may therefore be obtained with the choice $d\mathbf{l} = \hat{\mathbf{r}}\,dr$ to be

$$E = -\frac{dV}{dr} = -\frac{q}{4\pi\epsilon_0}\frac{d}{dr}\left(\frac{1}{r}\right) = \frac{q}{4\pi\epsilon_0}\frac{1}{r^2}$$

The calculation of the potential function V associated with a collection of charged particles proceeds in the same way as that for (21-9). The result is

$$V = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{q_i}{r_i} \qquad \textbf{(21-10)}$$

where q_i $(i = 1, 2, \ldots, N)$ is the charge of the ith particle and r_i is its distance from the field point. Thus, whereas in computing the electric field due to a collection of charged particles it is necessary to evaluate the vector sum of the electric fields due to each one, in calculating the associated potential function we need only add together the scalar potentials associated with each particle. From a practical point of view, the latter procedure is considerably simpler.

Note that in (21-9) the symbol r, which represents the distance from the particle to the field point, is an inherently positive quantity. In particular, then, the sign of the potential function for a charged particle is the same as the sign of the particle's charge. An analogous statement applies to the more general formula in (21-10).

Example 21-5 How much work is required to bring a particle of charge 6.0 μC from infinity to a point 10^{-2} meter from a fixed particle of charge -3.0 μC?

Solution According to (21-9), the potential V_0 at a distance of 10^{-2} meter from a particle of charge -3.0 μC is

$$V_0 = \frac{q}{4\pi\epsilon_0}\frac{1}{r} = \left(9.0 \times 10^9\,\frac{\text{N-m}^2}{\text{C}^2}\right) \times \frac{(-3.0 \times 10^{-6}\,\text{C})}{10^{-2}\,\text{m}}$$

$$= -2.7 \times 10^6\ \text{V}$$

where the minus sign is due to the fact that the particle has a negative charge. The work required to bring a particle of charge q from infinity to a point where the potential has the value V_0 is qV_0, according to (21-5). Using the known values for q and V we thus obtain

$$W = (6.0 \times 10^{-6}\,\text{C}) \times (-2.7 \times 10^{6}\,\text{V}) = -16\,\text{J}$$

The negative sign here signifies that the *electric field* has carried out positive work on the particle.

Example 21-6 Four particles each of charge q are located at the vertices of a square of side a, as in Figure 21-7. Calculate the potential at the center of the square.

Solution The diagonals of the square each have a length $a\sqrt{2}$, and hence the distance of each particle from the field point is $a\sqrt{2}/2$. Substituting this into (21-10), we obtain for V

$$V = \frac{1}{4\pi\epsilon_0}\sum \frac{q_i}{r_i} = \frac{1}{4\pi\epsilon_0} q \frac{4}{a\sqrt{2}/2}$$

$$= \frac{\sqrt{2}q}{\pi\epsilon_0 a}$$

What is the electric field at this field point?

Example 21-7 A charge q is placed on a conducting sphere of radius a. Show that (21-9) correctly gives the potential outside of the sphere ($r > a$) and calculate the potential V_0 of the sphere.

Solution By symmetry we know that the charge will distribute itself uniformly over the surface of the conducting sphere. It follows then by use of the result of Problem 38 of Chapter 20 that the electric field outside the sphere is the same as it would be if the entire charge q were concentrated at its center. Thus, for $r > a$, (21-9) correctly describes the potential due to the sphere.

Now, by the definition in (21-3), the potential function is differentiable, and hence it must be continuous. It follows that the potential V_0 of the surface of the sphere is

$$V_0 = \frac{q}{4\pi\epsilon_0}\frac{1}{a}$$

The fact that all points of the conducting sphere must be at the same potential was established in Example 21-4.

Figure 21-8 shows a plot of this potential. Note that even though the potential is continuous, its derivative—that is, the slope of this curve—is not continuous at $r = a$. This reflects the fact that the magnitude of the electric field vanishes inside the conductor and jumps discontinuously to the value $(q/4\pi\epsilon_0 a^2)$ just outside.

Example 21-8 Derive the formula

$$V = \frac{p}{4\pi\epsilon_0}\frac{z}{(x^2 + y^2 + z^2)^{3/2}}$$

for the potential of a dipole of moment p located at the origin and oriented along the positive z-axis.

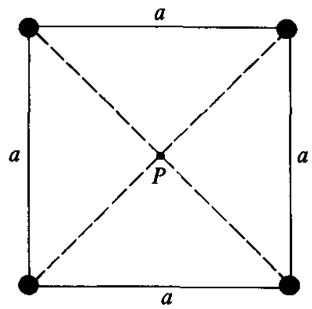


Figure 21-7

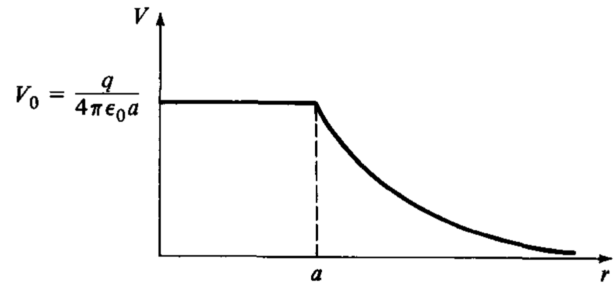


Figure 21-8

Solution Consider, in Figure 21-9, two particles of charges $+q$ and $-q$ located at the respective points $(0, 0, a)$ and $(0, 0, -a)$ in the given coordinate system. According to (21-10), the potential V for this charge configuration is

$$V = \frac{q}{4\pi\epsilon_0}\left\{\frac{1}{[x^2 + y^2 + (z - a)^2]^{1/2}} - \frac{1}{[x^2 + y^2 + (z + a)^2]^{1/2}}\right\} \tag{21-11}$$

since, for example, $[x^2 + y^2 + (z - a)^2]^{1/2}$ is the distance from the field point with coordinates (x, y, z) to the positive particle at $(0, 0, a)$.

Now, by definition, the dipole potential is obtained from the above formula by carrying out the limits $a \to 0$ and $q \to \infty$ in such a way that the product $p = 2aq$ remains finite. Making use of the formula

$$\frac{1}{(1 + \epsilon)^{1/2}} \cong 1 - \frac{1}{2}\epsilon \qquad [|\epsilon| \ll 1]$$

the first term in (21-11) becomes (in the limit as $a \to 0$)

$$\frac{1}{[x^2 + y^2 + (z - a)^2]^{1/2}} = \frac{1}{[r^2 - 2az + a^2]^{1/2}} = \frac{1}{r}\frac{1}{\left[1 + \dfrac{a^2 - 2az}{r^2}\right]^{1/2}} \cong \frac{1}{r}\left[1 + \frac{az}{r^2}\right] \tag{21-12}$$

where $r = (x^2 + y^2 + z^2)^{1/2}$. In a similar way the second term in (21-11) may be approximated by

$$-\frac{1}{[x^2 + y^2 + (z + a)^2]^{1/2}} \cong -\frac{1}{r}\left[1 - \frac{az}{r^2}\right] \tag{21-13}$$

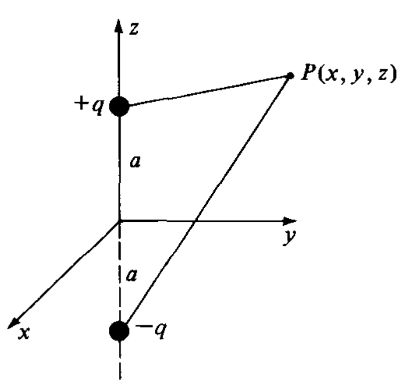


Figure 21-9

Substituting (21-12) and (21-13) into (21-11), we thus obtain

$$V=\frac{q}{4\pi\epsilon_0}\left\{\frac{1}{r}+\frac{az}{r^3}-\frac{1}{r}+\frac{az}{r^3}\right\}=\frac{2aqz}{4\pi\epsilon_0 r^3}=\frac{pz}{4\pi\epsilon_0 r^3}$$

where the final equality follows from the definition $p=2aq$.

21-6 The potential function for continuous charge distributions

In this section we illustrate the application of (21-10) to continuous charge distributions by reference to several examples. For a discussion of the physical significance of these distributions, see Section 20-4.

Example 21-9 Calculate the potential on the axis of and at a distance z from the plane of a circular loop of radius a carrying a charge per unit length λ (Figure 21-10).

Solution According to (21-9), the potential dV at the given field point due to the element of charge $\lambda a\,d\theta$ located at the angular displacement θ from a fixed reference line is

$$dV=\frac{\lambda a\,d\theta}{4\pi\epsilon_0}\frac{1}{[a^2+z^2]^{1/2}}\;.$$

Integrating over all values of θ from 0 to 2π we obtain for the potential V due to the entire ring

$$V=\int dV=\int_0^{2\pi}\frac{\lambda a\,d\theta}{4\pi\epsilon_0}\frac{1}{[a^2+z^2]^{1/2}}=\frac{\lambda a}{4\pi\epsilon_0(a^2+z^2)^{1/2}}\int_0^{2\pi}d\theta$$

$$=\frac{\lambda a}{2\epsilon_0}\frac{1}{(a^2+z^2)^{1/2}} \qquad \textbf{(21-14)}$$

It is left as an exercise to confirm by substitution into (21-7) that this formula yields the electric field in (20-7), as it must.

Example 21-10 A circular disk of radius R has a uniform charge per unit area σ. Calculate the potential at a point on the axis and at a distance z from the disk (Figure 21-11).

Solution As for the corresponding problem of the electric field calculation, imagine the disk as consisting of a series of concentric rings, a typical one of which has a radius r and thickness dr. The area of this ring is $2\pi r\,dr$ and thus it contains a total charge $2\pi\sigma r\,dr$. According to (21-14), the potential dV due to this ring is

$$dV=\frac{1}{4\pi\epsilon_0}2\pi\sigma r\,dr\frac{1}{(r^2+z^2)^{1/2}}$$

where we have made the substitution $a\to r$ since the radius of the loop in the present case is r. The potential V due to the entire disk is obtained by integration over all

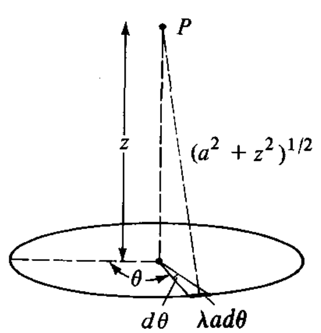


Figure 21-10

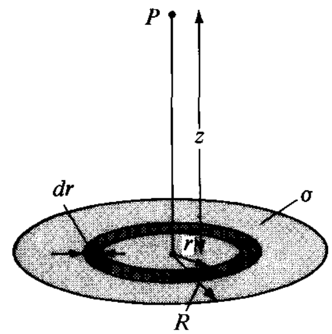


Figure 21-11

values of r from 0 to R. Thus we obtain

$$\begin{aligned} V &= \int dV = \int_0^R \frac{2\pi\sigma r\,dr}{4\pi\epsilon_0} \frac{1}{(r^2+z^2)^{1/2}} \\ &= \frac{\sigma}{2\epsilon_0} \int_0^R \frac{r\,dr}{(r^2+z^2)^{1/2}} = \frac{\sigma}{2\epsilon_0} (r^2+z^2)^{1/2} \Big|_0^R \\ &= \frac{\sigma}{2\epsilon_0} [(R^2+z^2)^{1/2} - \sqrt{z^2}] \end{aligned} \qquad \textbf{(21-15)}$$

Again, the consistency of this formula with (20-8) may be established by substitution into the third equation of (21-7).

Example 21-11 A sphere of radius a carries a uniform surface charge per unit area σ. Calculate the potential at a point P outside of the sphere by direct integration.

Solution Consider, in Figure 21-12, a circular element of the sphere of radius $a \sin\theta$ and of thickness $a\,d\theta$. The area of this element is $(2\pi a \sin\theta)(a\,d\theta)$ and hence it has a total charge $2\pi\sigma a^2 \sin\theta\,d\theta$. According to (21-14), the potential dV at P due to this charge is

$$\begin{aligned} dV &= \frac{2\pi\sigma a^2 \sin\theta\,d\theta}{4\pi\epsilon_0} \frac{1}{[(a\sin\theta)^2 + (r - a\cos\theta)^2]^{1/2}} \\ &= \frac{\sigma a^2 \sin\theta\,d\theta}{2\epsilon_0} \frac{1}{[a^2 + r^2 - 2ar\cos\theta]^{1/2}} \end{aligned}$$

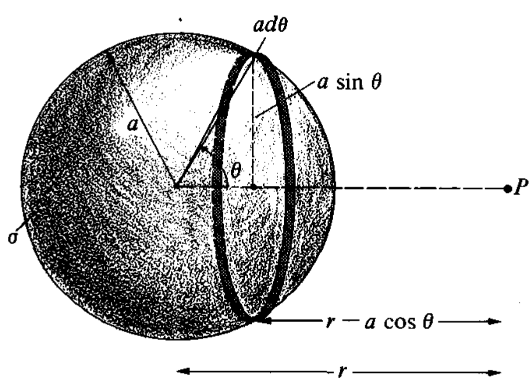


Figure 21-12

since the radius of the ring is $a \sin\theta$ and the distance from the center of the ring to the field point is $(r - a\cos\theta)$. The potential V due to the total sphere is obtained by integrating for all values of θ from 0 to π:

$$\begin{aligned}
V = \int dV &= \frac{\sigma a^2}{2\epsilon_0}\int_0^\pi \frac{\sin\theta\, d\theta}{[r^2 + a^2 - 2ar\cos\theta]^{1/2}} \\
&= \frac{\sigma a^2}{2\epsilon_0}\frac{[a^2 + r^2 - 2ar\cos\theta]^{1/2}}{ar}\bigg|_0^\pi \\
&= \frac{\sigma a^2}{2\epsilon_0 ar}\{\sqrt{(a+r)^2} - \sqrt{(r-a)^2}\} = \frac{\sigma a^2}{2\epsilon_0 ar}2a \\
&= \frac{Q}{4\pi\epsilon_0}\frac{1}{r}
\end{aligned} \tag{21-16}$$

where $Q = 4\pi a^2\sigma$ is the total charge on the sphere. This result is consistent with that found less directly in Example 21-7.

21-7 Equipotential surfaces and conductors

Consider a region of space in which there exists an electric field $\mathbf{E}$, with its associated potential function V. A surface S (be it a purely mathematical or an actual physical surface) is said to be an *equipotential surface* if the value of the potential at each point of the surface is the same. Thus, for a given potential function $V \equiv V(x, y, z)$ a given surface S is an equipotential surface at the value V_0 provided that all points (x, y, z) that satisfy the relation

$$V(x, y, z) = V_0$$

lie on this surface S and conversely all points of S satisfy this relation.

Consider, for example, the potential function associated with two particles of charges $+q$ and $-q$ and located in a given coordinate system at the respective points $(0, 0, a)$ and $(0, 0, -a)$. The potential function V is

$$V(x, y, z) = \frac{q}{4\pi\epsilon_0}\left\{\frac{1}{[x^2 + y^2 + (z-a)^2]^{1/2}} - \frac{1}{[x^2 + y^2 + (z+a)^2]^{1/2}}\right\}$$

Hence the points that lie on the equipotential surface characterized by the value V_0 satisfy the relation

$$V_0 = \frac{q}{4\pi\epsilon_0}\left\{\frac{1}{[x^2 + y^2 + (z-a)^2]^{1/2}} - \frac{1}{[x^2 + y^2 + (z+a)^2]^{1/2}}\right\}$$

In particular, the entire x-y plane defined by the equation $z = 0$ is an equipotential surface corresponding to the value $V_0 = 0$.

Similarly, since for a particle of charge q the potential function is

$$V = \frac{q}{4\pi\epsilon_0}\frac{1}{r}$$

it follows that the associated equipotential surfaces are the concentric spheres, r = constant, centered at the given particle. Specifically, the sphere of radius b is equipotential at the value V_0 given by

$$V_0 = \frac{q}{4\pi\epsilon_0}\frac{1}{b}$$

Figure 21-13 shows some of the equipotential surfaces (the dashed lines) and the associated electric field lines for these two systems. The vertical dashed line in Figure 21-13a represents the zero potential surface.

Equipotential surfaces have various interesting properties. Two of these, which follow directly from the definition of an equipotential surface are:

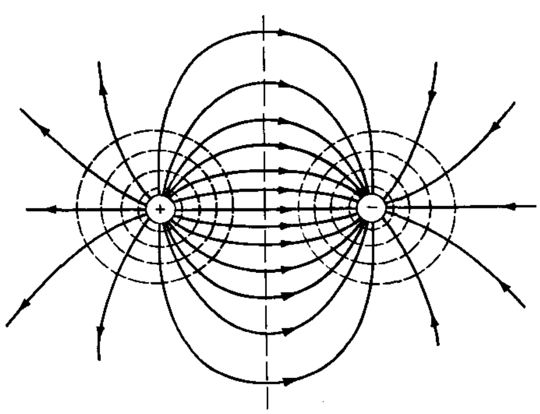
1. The work required to take a charged particle from one point on an equipotential surface to another point on this same surface is zero.
2. The electric field lines are everywhere perpendicular to the equipotential surfaces.

The first of these properties follows directly from (21-5). For if the points A and B lie on the same equipotential surface, then by definition $V_A = V_B$ and thus $W_{AB} = 0$. The second property follows from (21-3):

$$\mathbf{E} \cdot d\mathbf{l} = -dV$$

For if $d\mathbf{l}$ is a small displacement that lies entirely on an equipotential surface, then by definition of such a surface $dV = 0$, and hence $\mathbf{E} \cdot d\mathbf{l} = 0$. In other words, $\mathbf{E}$ must be perpendicular to $d\mathbf{l}$. Since by hypothesis $d\mathbf{l}$ lies entirely on the equipotential surface, it follows then that $\mathbf{E}$ must be perpendicular to this surface. This property that the electric field lines and the equipotential surfaces are mutually orthogonal is exemplified in Figure 21-13.

One of the very important applications of this notion of equipotential


(a)

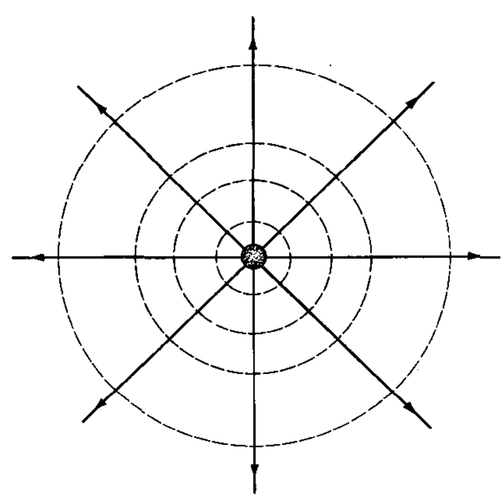

(b)

Figure 21-13

surfaces deals with conductors in an electric field. In Example 21-4 it was established that if a conductor is in an electric field, then the conducting surface must itself be an equipotential surface and have the same value as does the potential inside. Consider, in Figure 21-14a, a region of space where there exists an electric field and let the dashed surface S represent a closed equipotential surface with the value $V_0 = 0$. If, as in Figure 21-14b, we place a solid conductor of precisely the same shape as the surface S at its location and ground the conductor so that its potential is also zero, then the electric field outside of the conductor will be precisely the same as it was originally. In the interior of the conductor, on the other hand, the electric field must now vanish. Hence *if an equipotential surface is replaced by a conductor at the same potential as is the surface, then all external electric field lines remain unaltered.* However, those field lines that originally penetrated the equipotential surface will now terminate or originate on induced charge on the surface of the conductor.

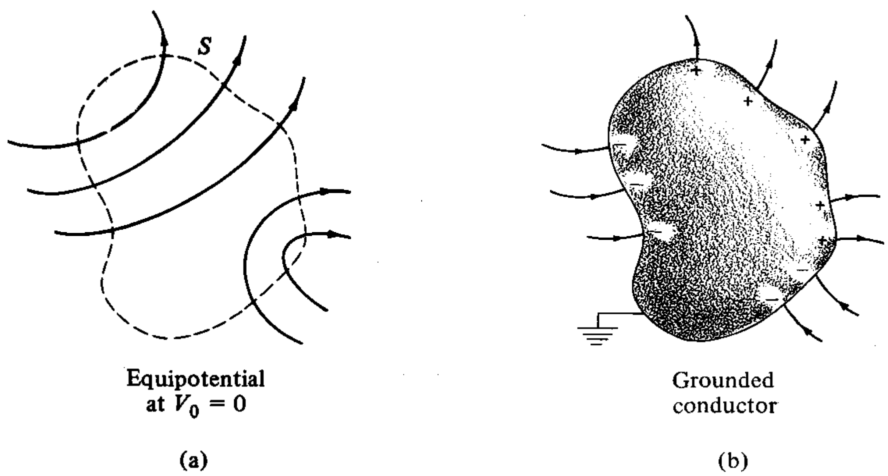


Figure 21-14

Example 21-12 A dipole of moment p_0 is located at the origin of a coordinate system and is oriented along its positive x-axis. Find the equations for the equipotential surfaces corresponding to $+V_0$ and $-V_0$.

Solution According to Example 21-8, the potential V for this dipole is

$$V(x, y, z) = \frac{p_0}{4\pi\epsilon_0} \frac{x}{(x^2 + y^2 + z^2)^{3/2}}$$

Thus the $V = +V_0$ surface satisfies the equation

$$(x^2 + y^2 + z^2)^{3/2} = \alpha x$$

with $\alpha = p_0/4\pi\epsilon_0 V_0$, and the $V = -V_0$ surface satisfies

$$(x^2 + y^2 + z^2)^{3/2} = -\alpha x$$

Transforming to polar coordinates, $r = (x^2 + y^2)^{1/2}$ and $\cos\theta = x/r$, we obtain the

projection of these two surfaces in the x-y plane:

$$r^2 = \pm\alpha \cos\theta$$

and these are plotted in Figure 21-15. The complete surface is obtained by rotating this figure about the x-axis.

Example 21-13 Two infinite, conducting planes are parallel, at a separation distance d and carry the respective uniform surface charge densities $+\sigma$ and $-\sigma$. Suppose the positively charged plate is grounded.

(a) Calculate the potential of the negatively charged plate.

(b) What is the potential V_0 of the equipotential surface at a distance x ($<d$) away from the positive plate?

Solution As shown in Figure 21-16, the field lines are uniform and perpendicular to the plates. They originate on the positively charged plate and terminate on the negative charge on the other plate. Since the plates are infinite in extent, the electric field is confined to the region between the plates. The equipotential surfaces must be perpendicular to the field lines and are thus planes parallel to the plates. These equipotential surfaces are designated by the dashed lines in the figure.

(a) According to (20-16), the magnitude of the electric field immediately to the right of the positive plate and to the left of the negative plate is

$$E = \frac{\sigma}{\epsilon_0}$$

and thus, according to the above argument, σ/ϵ_0 is the electric field strength everywhere between the plates. Hence the work required to take a unit charge from the positive to the negative plate is $-\sigma d/\epsilon_0$. Making use of (21-5) and the fact that the left-hand plate is grounded, we conclude then that the potential of the negatively charged plate is $-\sigma d/\epsilon_0$.

(b) By use of the same arguments as above, the potential $V(x)$ at a distance x ($<d$) from the positive plate is found to be

$$V(x) = -\frac{\sigma}{\epsilon_0}x$$

This time the work required to take a unit charge from the positive plate a distance x toward the negative one is $-\sigma x/\epsilon_0$.

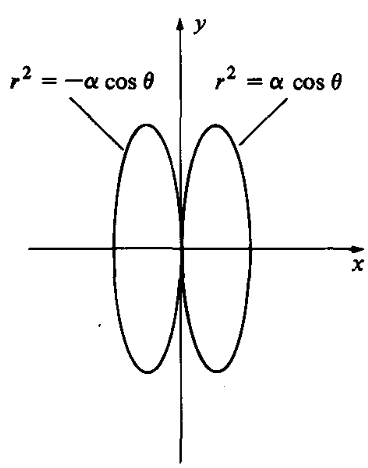


Figure 21-15

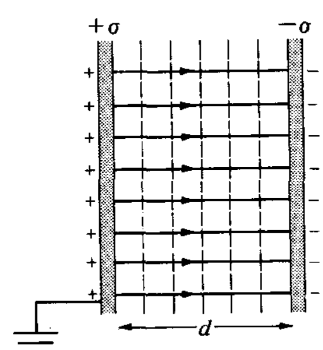


Figure 21-16

21-8 Energy in the electrostatic field

The notion of energy plays a very important and often central role in many branches of physical science. The purpose of this section is to examine this concept as it relates to the electrostatic field.

Consider first the case of two particles of charges q_1 and q_2 separated by a distance r_{12}. The electrostatic energy U_2 associated with this two-particle system is numerically equal to the work W_2 required to bring them from a state of infinite separation to the given configuration. According to (21-9), the potential function V associated with q_1 at a point a distance r_{12} away is

$$V = \frac{q_1}{4\pi\epsilon_0}\frac{1}{r_{12}}$$

But the work W_2 required to bring q_2 from a state of infinite separation to within a distance r_{12} of q_1 is $q_2 V$. Hence, equating this to the energy U_2 of this two-particle system, we find that

$$U_2 = \frac{q_1 q_2}{4\pi\epsilon_0}\frac{1}{r_{12}} \tag{21-17}$$

Consistent with the facts that charged particles of the same sign repel and those of unlike sign attract, U_2 is positive in the former case and negative in the latter. In other words, if the force between the particles is repulsive, then consistent with (21-17), positive work must be carried out on the system, and $U_2 > 0$. For the case $q_1 q_2 < 0$ the field does work on the agent since the electric force is attractive, and hence $U_2 < 0$ in this case.

The calculation of the electrostatic energy U_3 for three charged particles proceeds along similar lines. This time, U_3 is the sum of the work U_2 required to bring q_1 and q_2 to their final positions, plus that required to bring q_3 to its final position at the point P once q_1 and q_2 are in place; see Figure 21-17. The value of the potential V at the point P due to q_1 and q_2 is, according to (21-10),

$$V = \frac{1}{4\pi\epsilon_0}\left[\frac{q_1}{r_{13}} + \frac{q_2}{r_{23}}\right]$$

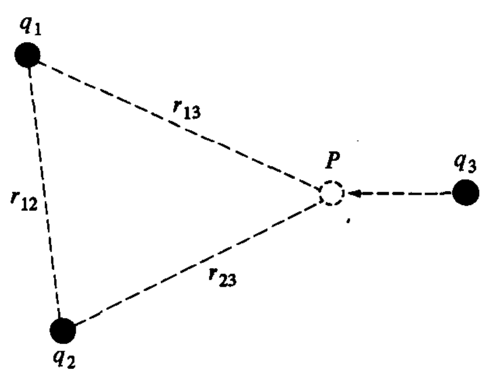


Figure 21-17

where r_{13} and r_{23} are the respective distances of P from q_1 and q_2. The work required to bring in q_3 is q_3V, and adding this to U_2 we obtain

$$U_3 = \frac{1}{4\pi\epsilon_0}\left[\frac{q_1q_2}{r_{12}} + \frac{q_1q_3}{r_{13}} + \frac{q_2q_3}{r_{23}}\right] \tag{21-18}$$

Note the symmetry of this formula under permutation of the indices. This is as it should be, since the order in which the particles are brought to the final configuration should have no physical significance. The corresponding formula for more than two particles should be evident by analogy to (21-18).

Example 21-14 Three particles, each of charge q, are at the vertices of an equilateral triangle of side a. Calculate:

(a) The electrostatic energy of this configuration.

(b) The change in this energy if the particles are pulled apart until the sides of the triangle have a length $2a$.

Solution

(a) In the notation of (21-18), the parameter values are $q_1 = q_2 = q_3 = q$ and $r_{12} = r_{23} = r_{13} = a$. Thus we find that

$$U = \frac{1}{4\pi\epsilon_0}\left(\frac{q_1q_2}{r_{12}} + \frac{q_1q_3}{r_{13}} + \frac{q_2q_3}{r_{23}}\right) = \frac{3q^2}{4\pi\epsilon_0}\frac{1}{a}$$

(b) By use of this result, it follows that if the sides of the triangle were of length $2a$, then the energy would be

$$\frac{3q^2}{4\pi\epsilon_0}\frac{1}{2a}$$

Thus the energy change ΔU or the work required to go from the former to the latter configuration is

$$\Delta U = \frac{3q^2}{4\pi\epsilon_0}\left(\frac{1}{2a} - \frac{1}{a}\right) = -\frac{3q^2}{8\pi\epsilon_0 a}$$

The fact that $\Delta U < 0$ means that the energy of the electrostatic field has decreased. This is consistent with the fact that the external agent must carry out *negative* work to bring about this configuration change.

Example 21-15 A spherical conductor of radius a has a total charge Q. What is the energy U of this system?

Solution To obtain U, let us calculate the work W required to bring the total charge Q from infinity to the surface of the sphere in successive, infinitesimal amounts. Suppose at some stage an amount of charge q has been brought up to the sphere. The work dW required to bring up an additional infinitesimal amount dq is

$$dW = V(q)\,dq = \frac{q}{4\pi\epsilon_0}\frac{1}{a}\,dq$$

where $V(q)$ is the potential of the sphere when it has charge q, and the second equality follows from (21-16) for $r = a$. The total work W and thus the electrostatic energy is obtained by integrating over all values of q from 0 to Q. The result is:

$$U = W = \int dW = \frac{1}{4\pi\epsilon_0 a}\int_0^Q q\,dq = \frac{Q^2}{8\pi\epsilon_0 a} \tag{21-19}$$

21-9 Spherically and cylindrically symmetric charge distributions

Although it is possible, in principle, to use the methods of Section 21-6 to derive formulas for the potential function associated with any continuous distribution of charge, the integrals involved are generally much too complex for the results to be practically useful. However, for the special case of charge distributions having spherical or cylindrical symmetry there is an alternate and practical method available. For these cases the electric field $\mathbf{E}$ points along the radial direction and varies only with the radial distance r. Hence the potential V is determined by

$$E = -\frac{dV}{dr} \tag{21-20}$$

the validity of which follows from (21-3) if we select for $d\mathbf{l}$ a small displacement dr along the radial direction. Note that the symbol r in this context is used in two senses. For situations having spherical symmetry it denotes the radial distance from the origin, whereas for cases of cylindrical symmetry it represents the radial distance from the axis.

To illustrate the usage of (21-20) let us consider two examples involving spherical symmetry. Applications to cases of cylindrical symmetry will be found in the problems.

Example 21-16 Find the potential function associated with a uniformly charged sphere of radius a and charge density ρ_0.

Solution Substituting (20-20) and (20-21) into (21-20), and making use of the fact that the total charge Q_0 of the sphere is $4\pi\rho_0 a^3/3$, we find that

$$\frac{dV}{dr} = -E = -\frac{\rho_0}{3\epsilon_0}\begin{Bmatrix} a^3/r^2 & \quad r \geqslant a \\ r & \quad r \leqslant a \end{Bmatrix}$$

Hence since $r = d(r^2/2)/dr$ and $d(-r^{-1})/dr = 1/r^2$, this may be integrated to

$$V(r) = -\frac{\rho_0}{3\epsilon_0}\begin{Bmatrix} c_1 - a^3/r & \quad r \geqslant a \\ c_2 + r^2/2 & \quad r \leqslant a \end{Bmatrix}$$

with c_1 and c_2 constants of integration. The boundary condition that $V(\infty)=0$ determines the value $c_1=0$, while the condition that $V(r)$ be continuous at $r=a$ determines the second constant c_2 to be $-3a^2/2$. Hence we obtain the final formula

$$V(r)=\frac{\rho_0}{3\epsilon_0}\begin{cases} a^3/r & r \geqslant a \\ \dfrac{3a^2}{2}-\dfrac{r^2}{2} & r \leqslant a \end{cases} \tag{21-21}$$

Example 21-17 Two concentric, conducting spherical shells of radii a and b have charges $+q$ and $-q$, respectively (Figure 21-18). Calculate the potential $V(a)$ of the smaller sphere assuming that the outer one is grounded so that $V(b)=0$.

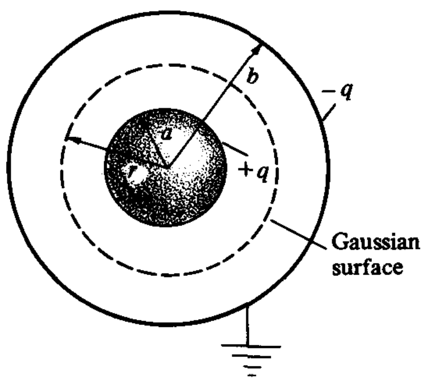


Figure 21-18

Solution Applying Gauss' law to a spherical surface of radius r $(a<r<b)$ we find, as in Example 20-13, that the electric field $E(r)$ $(a<r<b)$ is radial and is the same as that of a point charge q. Thus (21-20) becomes in this case

$$\frac{dV}{dr}=-E=-\frac{q}{4\pi\epsilon_0}\frac{1}{r^2} \qquad a<r<b$$

Integrating both sides, from $r=a$ to $r=b$, we find that

$$V(b)-V(a)=-V(a)=-\frac{q}{4\pi\epsilon_0}\int_a^b\frac{dr}{r^2}=-\frac{q}{4\pi\epsilon_0}\left[-\frac{1}{r}\right]\Bigg|_a^b$$
$$=\frac{q}{4\pi\epsilon_0}\left[\frac{1}{b}-\frac{1}{a}\right]$$

The fact that $V(a)$ is positive (for $q>0$) indicates that the outer sphere with its negative charge is at the lower potential. As noted previously, the direction of the electric field is toward decreasing values for potential.

21-10 Summary of important formulas

The work W_{AB} carried out by an agent in the presence of an electric field $\mathbf{E}$ in slowly transporting a particle of charge q_0 from A to B is

$$W_{AB}=-q_0\int_A^B \mathbf{E}\cdot d\mathbf{l} \tag{21-2}$$

and this is the same for *all paths* connecting the two points A and B. The potential function V associated with $\mathbf{E}$ is defined so that the change dV between two points at a relative infinitesimal displacement $d\mathbf{l}$ is

$$dV = -\mathbf{E} \cdot d\mathbf{l} \tag{21-3}$$

Substitution into (21-2) yields the relation between work and potential difference

$$W_{AB} = q_0(V_B - V_A) \tag{21-5}$$

The potential function V at a distance r from an isolated particle of charge q is

$$V = \frac{q}{4\pi\epsilon_0}\frac{1}{r} \tag{21-9}$$

and for a collection of N particles of charges $q_1, q_2, \ldots, q_n$, the potential V is

$$V = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{q_i}{r_i} \tag{21-10}$$

where r_i is the distance of the ith particle from the field point.

The energy U_2 associated with a two-particle configuration is the work required to bring them from a state of infinite separation to the given state. It has the value

$$U_2 = \frac{q_1 q_2}{4\pi\epsilon_0}\frac{1}{r_{12}} \tag{21-17}$$

where r_{12} is the distance between the two particles.

QUESTIONS

1. Define or describe briefly what is meant by (a) conservative field; (b) potential difference; (c) equipotential surface; and (d) electrostatic energy.
2. Consider two potential functions V_1 and V_2, which characterize the *same* charge distribution. In what way or ways can they differ from each other? Assuming that $V_1 \neq V_2$, explain why they are physically indistinguishable.
3. The potential of the earth is customarily assigned the value of zero volts. What effect would a different assignment, say 10 volts, have on measured values of potentials? What effect would there be on measurements of potential differences?
4. If the force between two charged particles varied as $r^{-2-\epsilon}$ ($\epsilon > 0$) instead of r^{-2}, as it actually does, would a potential function for the electrostatic field exist? Would Gauss' law be valid in this circumstance?
5. A positive charge q is transferred from conductor A to conductor B. Assuming A and B are originally electrically neutral, which of them will be at the higher potential? In what way would your answer be different if A and B originally carried a charge?
6. Explain in qualitative terms why the line integral of a conservative field around a *closed* path must vanish. Is this equivalent to the text's definition of such a field?

7. As you go along the direction of an electric field line, do you go to regions of higher or of lower potential? Justify your answer by use of (21-3).
8. A proton is initially at rest in an electric field. Explain why it will go to a region of lower potential. What would happen to an electron under the same circumstance?
9. Consider the physical situation in Example 21-1. If the particle goes from B to A in Figure 21-2, how much work is done by the external agent? (Make use of the result obtained in the example.) How much work is carried out by the electric field? Which of these two works must be positive? Assume that σ is positive.
10. How much work would be carried out for the system in Figure 21-2 if the particle were taken from A to B then to D and finally back to A? Is it necessary for your answer that this path be entirely along the linear paths in the figure?
11. Show that it is possible for the potential to be zero at a point where the electric field does *not* vanish. Is it possible for the electric field to vanish at a point where the potential function does not?
12. Assuming that the potential at infinity vanishes, in Example 21-6 we established that the potential at the center of the square in Figure 21-7 is nonzero. Nevertheless, the electric field vanishes at this point. Is this a contradiction? Explain.
13. Explain why the term $\sqrt{z^2}$ in (21-15) cannot be replaced by z. (*Hint*: Compare the direction in which the electric field would point on opposite sides of the disk if this replacement were made.)
14. Why must the potential be a continuous function of position? (*Hint*: Make use of (21-3) and show that this relation cannot be satisfied for finite **E** at a point where the potential is discontinuous.)
15. What can you say about the electric field in a region where the potential is constant? What can you say about the potential in a region of constant electric field?
16. Is it possible for surfaces, characterized by different values for potential, to intersect? Can two such surfaces be tangent to each other? Explain.
17. Must all equipotential surfaces be closed? If your answer is no, give an example of a physical situation for which an equipotential surface is open. If your answer is yes, explain.
18. What simplifications (if any) occur in using Gauss' law with a surface that is equipotential? Need the electric field have the same value everywhere on such a surface?
19. Consider an infinitely long cylinder that contains a uniform charge. Describe the equipotential surfaces.
20. Why must every point in the interior of a conducting shell be at the same potential as the shell is? Assume that there is no charge in the cavity, but that the conductor itself carries a charge.
21. A spherical conductor of radius a has a charge Q. What are the equipotential surfaces inside and outside the conductor? What is the potential of the conductor?
22. If two particles of the same sign of charge are brought closer together, does the electrostatic energy increase or decrease? What happens if they are of the opposite sign?
23. A particle of charge q is brought near an uncharged conductor. Does the electrostatic energy increase or decrease in this process? Is the sign of q relevant to your answer?

PROBLEMS

1. A particle of charge 6.0 μC is in a uniform field $\mathbf{E}$ of strength 200 V/m. How much work must be supplied by an external agent to move the particle a distance of 0.5 meter:
   (a) Parallel to the field?
   (b) Antiparallel to the field?
   (c) At right angles to the field?
   (d) Along a line inclined at 30° to the field?
2. By use of the fact that for a conservative field $\mathbf{E}$ the line integral

$$\int_A^B \mathbf{E} \cdot d\mathbf{l}$$

   is independent of the path connecting the points A and B, show that the line integral of $\mathbf{E}$ around a closed path is zero. Derive the converse of this result.
3. An infinite nonconducting plane of charge has a density $\sigma = 10^{-6}$ C/m^2. (a) What is the electric field strength? (b) Describe the equipotential surfaces. (c) If the plane has zero potential, what is the potential of a point 0.2 meter from the plane?
4. A uniform electric field has a strength of 2.0×10^5 V/m. (a) How far apart are two equipotential surfaces whose potential difference is 2.0 volts? (b) What is the potential difference between two equipotential surfaces separated by a distance of 0.4 meter?
5. Two very large metal plates are separated by a distance of 1 cm and are maintained at a potential difference of 100 volts. (a) What is the strength of the uniform electric field in the region between the plates? (b) How much work is required to take a particle of charge +6.0 μC from the plate at the higher potential to the other?
6. What is the strength of the potential at a distance of 2.0×10^{-4} meter from a particle of charge 3.0 μC? At a distance of 2.0 km? Assume in each case that the potential at infinity vanishes.
7. Figure 21-19 shows a particle of charge $+3q$ at a distance d from a particle of charge $-q$. (a) What is the potential at the point P at a distance a from $3q$? (b) For what value of a will the potential at P vanish?

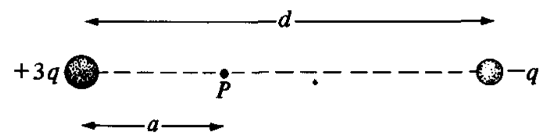


Figure 21-19

8. For the charge configuration in Figure 21-19 calculate:
   (a) The potential at the midpoint of the line joining the two particles.
   (b) The potential at a distance $d/2$ along the perpendicular bisector of the line joining the particles.
   (c) The work required to take a particle of charge q_0 from the point in (a) to the point in (b).
9. The electron and the proton in a hydrogen atom are separated by a distance of 0.53×10^{-10} meter. Assuming the charge values are $\mp 1.6 \times 10^{-19}$ coulomb, calculate the potential at:
   (a) The midpoint of the line joining them.
   (b) A distance of 0.1 mm along their perpendicular bisector.
   (c) How much work is required to bring a second electron from the point in (b) to that in (a)?
10. A particle of charge +1.0 μC is located at the point (0, 0, 1) in a cer-

tain coordinate system. A second particle has charge $-2.0\ \mu C$ and is located at $(0, 0, -1)$. Assuming that all lengths are in meters, calculate the value of the potential at the following points: **(a)** $(0, 0, 0)$; **(b)** $(0, 0, 2)$; **(c)** $(1, 1, 0)$.

11. How much work is required to take a particle of charge $-3.0\ \mu C$ from the initial point $(2, 2, 2)$ to each of the three points where the potential has been evaluated in Problem 10?

12. Assume in Figure 21-7 that each particle has a charge q. **(a)** What is the potential at the midpoint of the upper side of the square? **(b)** How much work is required to take a particle of charge $-q$ from the midpoint of the upper side to the midpoint of the lower side? **(c)** Explain your answer to (b) in physical terms.

13. For the charge configuration in Figure 21-7 calculate: **(a)** The potential at a perpendicular distance y above the plane of the square and vertically above its center. **(b)** The electric field at the point in (a) by use of (21-7).

14. Three particles of charges q, q, and $-3q$ are at the vertices of an equilateral triangle of side a. Calculate the potential at **(a)** the midpoint of a side whose ends carry the same charge and **(b)** a point in the plane of the triangle and equidistant from the three vertices.

15. Calculate everywhere the electric field **E** associated with the potential function

$$V = -\alpha(x^2 + y^2 - 2z^2)$$

with α a positive constant.

16. The components of the electric field in a certain coordinate system are (in SI units) $E_x = -10$; $E_y = 3y$; and $E_z = 0$. Calculate a potential function for this field.

17. Show that the components of **E** must satisfy

$$\frac{\partial E_x}{\partial y} = \frac{\partial E_y}{\partial x}$$

and two similar relations obtained by cyclic permutation of the indices. (*Hint*: Express the components of the field in terms of the derivatives of the potential.)

18. A dipole of moment p is oriented along the positive x-axis in a certain coordinate system.

(a) Show that in the x-y plane the potential V may be written

$$V = \frac{p}{4\pi\epsilon_0}\frac{\cos\theta}{r^2}$$

where $r = (x^2 + y^2)^{1/2}$ and $\cos\theta = x/r$ are polar coordinates.

(b) Show by use of (21-3) that the components of **E** along the direction of increasing r and of increasing θ, namely E_r and E_θ, respectively, are given by

$$E_r = -\frac{\partial V}{\partial r} \qquad E_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta}$$

19. A circular disk of radius R_0 carries charge per unit area σ. How much work is required to take a particle of charge q_0 from a point on the axis of the disk and at a distance z from its plane to: **(a)** The point on the axis at a distance z on the other side of the disk? **(b)** The center of the disk?

20. By use of (21-15) calculate the electric field at any point on the axis of a disk of radius R and carrying a uniform charge per unit area σ.

***21.** A line charge of length $2l$ has charge per unit length λ. Show by direct integration that the potential at a distance r from the line charge and on its perpendicular bisector is

$$V = \frac{\lambda}{4\pi\epsilon_0}\ln\frac{[l + (l^2 + r^2)^{1/2}]}{[-l + (l^2 + r^2)^{1/2}]}$$

22. For the physical system in Problem 21: (a) calculate the electric field along the perpendicular bisector and (b) show that, for an infinitely long line charge, that is, for $r/l \to 0$, your formula for E reduces to that in (20-6).

23. A nucleus of atomic weight A contains Z protons and may be thought of as a uniform sphere of charge of radius r_0 given by

$$r_0 = 1.1\, A^{1/3} \times 10^{-15}\ \text{m}$$

(a) In terms of A, Z, and the proton charge e, what is the potential at any point outside of the nucleus?
(b) Calculate the potential at any point inside the nucleus.
(c) How much work is required to take an electron from the surface of the nucleus to a distance of 10^{-10} meter from its center? Assume that $A = 235$ and $Z = 92$.

24. A charge of 10^{-7} coulomb is placed on a conducting sphere having a radius of 0.5 meter. (a) What is the potential of the sphere? (b) What is the value of the potential at a distance of 2 meters from the center of the sphere? (c) What is the potential at any point inside the sphere?

25. Figure 21-20 shows two infinite conducting planes, which are parallel to each other and carry uniform surface charge densities $+\sigma$ and $-\sigma$, respectively. Assume that the positive plate is grounded and that an infinite planar conductor of thickness b is introduced at a distance a from the positively charged plate.
(a) Show that the potential at a distance x ($<a$) from the positive plate is

$$V(x) = -\frac{\sigma}{\epsilon_0} x$$

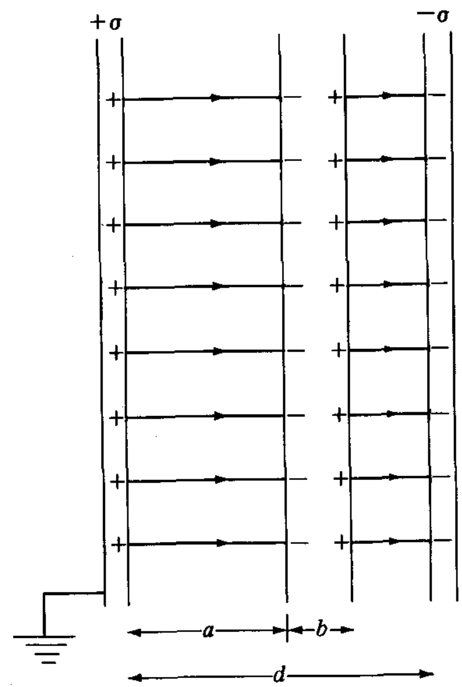


Figure 21-20

(b) What is the potential of the conducting slab?
(c) Calculate the potential of the negatively charged plate and contrast your answer with the corresponding one in Example 21-13.

26. Two particles of charges $-qa/b$ and q are located at the points with coordinates $(a^2/b, 0, 0)$ and $(b, 0, 0)$, respectively, with a and b ($>a$) certain positive distances. (a) Show that a sphere of radius a centered at the origin is an equipotential surface. (b) What value for the potential is associated with this surface?

27. Repeat both parts of Problem 26, but suppose the presence this time of a third particle of charge Q_0 located at the origin.

28. How much work is required to bring the three charges Q_0, $-qa/b$, and q in Problem 27 from a state of infinite separation to the given configuration?

29. What is the electrostatic energy as-

sociated with the charge configuration in Figure 21-7?

30. (a) Calculate the electrostatic energy associated with the two particles in Figure 21-19. (b) What would this energy be if a third particle of charge $-q$ were located at the point P in the figure?

31. Calculate the electrostatic energy associated with a hydrogen atom assuming that the electron and proton are at a separation of 0.53×10^{-10} meter.

32. Two particles, each of charge 2.0 μC, are at a separation distance of 0.4 meter. (a) What is the energy associated with this configuration? (b) Where can a third particle of charge $-3.0\ \mu$C be placed so that the electrostatic energy of the resulting configuration will be zero? Is this point unique?

33. A regular tetrahedron of side a has a charged particle at each of its four vertices. If three of these have a charge $+q$ and the fourth a charge $-3q$, calculate the electrostatic energy of this configuration. How much work is required to double the sides of the tetrahedron?

34. Consider a nonconducting sphere of radius r, which carries a uniform charge per unit volume ρ_0.
   (a) Show by use of (21-21) that the potential V at the surface of this sphere is

$$V = \frac{\rho_0}{3\epsilon_0} r^2$$

   (b) Show that the change in electrostatic energy dU, on adding to this sphere a spherical shell of radius r and thickness dr and thus carrying a total charge $4\pi r^2\, dr\, \rho_0$, is

$$dU = \frac{4\pi}{3\epsilon_0} \rho_0^2 r^4\, dr$$

   (c) By making use of (b), prove that the energy U associated with a sphere of radius a and carrying a uniform charge density ρ_0 is

$$U = \frac{3}{5} \frac{1}{4\pi\epsilon_0} \frac{Q^2}{a}$$

   where

$$Q = \frac{4\pi}{3} a^3 \rho_0$$

   is the total charge on the sphere.

35. (a) Show by use of the result of Problem 34 that the energy $U(A, Z)$ of a nucleus of atomic weight A and charge Ze, assuming that it is a sphere of uniform charge and of radius $a = 1.1 \times 10^{-15} A^{1/3}$ meter, is

$$U(A, Z) = \frac{3}{5} \frac{e^2}{4\pi\epsilon_0} \frac{Z^2}{1.1 \times 10^{-15} A^{1/3}}$$

   Why do you suppose that this formula is normally written with the Z^2 factor replaced by $Z(Z-1)$?
   (b) What is the "potential gradient"—that is, the electric field strength—at the surface of the nucleus $^{235}_{92}U$?

36. Using (21-19) calculate the electrostatic energy stored in a conducting sphere of radius 2.0 cm that contains a charge of 10 μC.

37. Making use of the result of Problem 34 calculate the "electrostatic energy" of a proton assuming it to be a uniformly charged sphere of radius 3.0×10^{-16} meter (see Figure 19-19). Compare this with the proton's "rest energy," Mc^2 (c = speed of light $\cong 3 \times 10^8$ m/s).

38. A particle of charge q and mass m undergoes motion in an electrostatic field $\mathbf{E}$ with associated potential function V. Show that the energy $\mathscr{E}$ of the particle

$$\mathscr{E} = \frac{1}{2} mv^2 + qV$$

where $\mathbf{v}$ is the velocity of the particle, is constant in time. (*Hint*: Calculate $d\mathscr{E}/dt$ and show by use of Newton's law and (21-7) that it vanishes.)

39. By use of the result of Problem 38, calculate the potential difference through which an electron travels if: **(a)** Its velocity drops from 2.0×10^4 m/s to zero. **(b)** Its velocity rises from 2.0×10^4 m/s to 3.0×10^6 m/s.

40. The unit of the "electron volt" (eV) is an energy unit that represents the gain or loss of kinetic energy of a particle of charge $\pm1.6\times10^{-19}$ coulomb when it goes through a potential drop of 1 volt. Find the relation between the electron volt and the joule.

41. A proton starts from rest, travels in a uniform field for a distance of 0.2 mm, and acquires a final velocity of 3.0×10^3 m/s. **(a)** Through what potential difference has it gone? **(b)** What is the strength of the electric field?

42. In an electrostatic accelerator doubly ionized helium atoms or alpha particles are accelerated from rest through a potential difference of 2.0×10^6 volts. Calculate the final velocity of the α particles.

43. A sphere of radius a contains a uniform charge density ρ_0 (>0) and is surrounded by a concentric grounded conductor of radius b (Figure 21-21). Calculate: **(a)** the potential for $r \leqslant a$; **(b)** the potential for $b \leqslant r \leqslant a$; and **(c)** the charge density σ on the surface of the conductor.

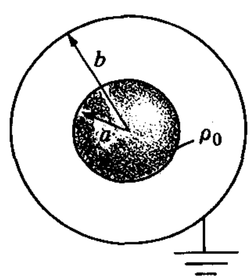


Figure 21-21

44. A particle of charge q is at the center of a spherical conducting shell of inner radius a and outer radius b ($>a$), which in turn is inside a concentric grounded spherical conductor of radius c ($>b$). Calculate:
(a) the potential for $r \leqslant a$;
(b) the potential for $a \leqslant r \leqslant b$;
(c) the potential for $b \leqslant r \leqslant c$.

***45.** A spherical hole of radius a is cut out of the center of a sphere of radius b ($>a$) and containing a uniform charge per unit volume ρ_0. Show that the potential in the hole is constant and determine the value of this constant assuming that $V = 0$ at infinity.

46. A conducting sphere of radius a carries a charge q.
(a) What is the potential of the sphere?
(b) What is the electric field strength at the surface of the sphere?
(c) The air surrounding such a sphere will break down (that is, a spark will be generated) if the electric field reaches a strength $\cong 3\times10^6$ V/m. How much charge can be placed on a sphere with a radius of 0.2 meter without causing a breakdown?

47. An infinite cylinder of radius a has a uniform charge per unit volume ρ_0. Show that the potential at a distance r from the axis of the cylinder is

$$V(r) = \frac{\rho_0}{2\epsilon_0} \times \begin{Bmatrix} -r^2/2 & r \leqslant a \\ -\dfrac{a^2}{2} - a^2 \ln\dfrac{r}{a} & r \geqslant a \end{Bmatrix}$$

provided $V(0) = 0$. Explain in physical terms why we *cannot* select $V(\infty) = 0$ in this case.

48. Consider two coaxial cylindrical conducting shells of radii a and b $(>a)$. Suppose that the outer one is grounded and the inner one has a charge per unit area σ.

(a) Calculate the electric field everywhere.

(b) What is the potential of the inner cylinder?

(c) Calculate the charge density on the outer cylinder.

22 Dielectric materials and capacitors

The beauty of electricity... is not that the power is mysterious and unexpected... but that it is under law and that the taught intellect can now even govern it largely.

MICHAEL FARADAY (1791-1867)

22-1 Introduction

In our studies of the effects of electric charge on matter, up to this point we have been concerned mainly with conductors. The purpose of this chapter is to discuss the electric properties of a second class of materials, which are known as *insulators* or *dielectrics.*

A solid conductor, it will be recalled, may be visualized as a rigid lattice of positive ions, throughout which is interspersed a compensating negative charge of highly mobile electrons. When placed into an electric field, these electrons distribute themselves in such a way that the resultant electric field vanishes throughout the conductor.

By contrast, in a dielectric each electron is, as a rule, tightly bound to its parent ion, and thus undergoes very little translational motion even in the presence of an external electric field. However, any externally applied electric field will, in general, bring about some degree of charge separation between each electron and its associated ion, and in this way an originally neutral atom becomes a small electric dipole. As will be seen below, it is

because of the existence of these dipoles that a dielectric acquires its distinctive electric properties when placed into an electric field.

In connection with a study of dielectrics it is convenient to introduce first the notion of *capacitance.* Accordingly, the first several sections of this chapter are concerned with a discussion of capacitors and with the derivation of some of their important properties. The remainder of the chapter is then devoted to an analysis of the macroscopic, electric properties of dielectrics using capacitors as a tool.

22-2 Capacitors and capacitance

Consider, in Figure 22-1, two isolated conductors A and B and suppose that a charge q (>0) has been taken from one of these—originally neutral—conductors and placed on the other. The conductors will then have the charges q and $-q$, respectively, and as shown in the figure, electric-field lines will originate on the surface of the positively charged conductor and terminate on the surface of the other. A configuration of two conductors of this type is known as a *capacitor* and the conductors themselves are often referred to as its *plates.* It is customary to assume that $q > 0$, so the symbol q represents the net charge on the positive plate of the capacitor. The charge on the other plate will always be $-q$.

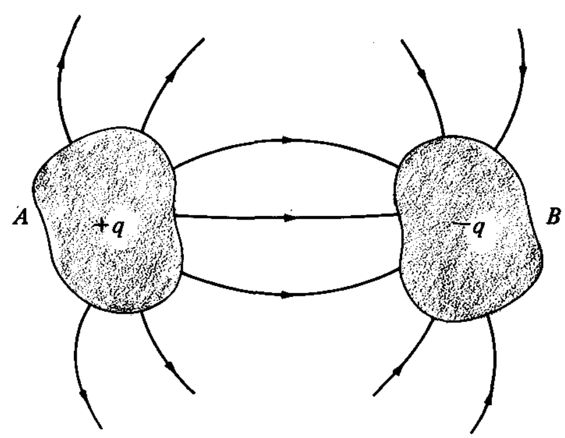


Figure 22-1

Figure 22-2 shows a simple way for charging a capacitor. Suppose that the two conductors A and B are originally electrically neutral and are connected by a conducting wire so that they are at the same potential. If a positive electric charge is brought near, say, conductor B, then some electrons will flow along the conducting wire from A to B. As shown in Figure 22-2b, at equilibrium A will have a certain positive charge q and its partner, B, will have a compensating charge $-q$ of the opposite sign. At this point they are still at the same potential. The remaining steps then involve disconnecting the wire joining A and B and finally removing the external

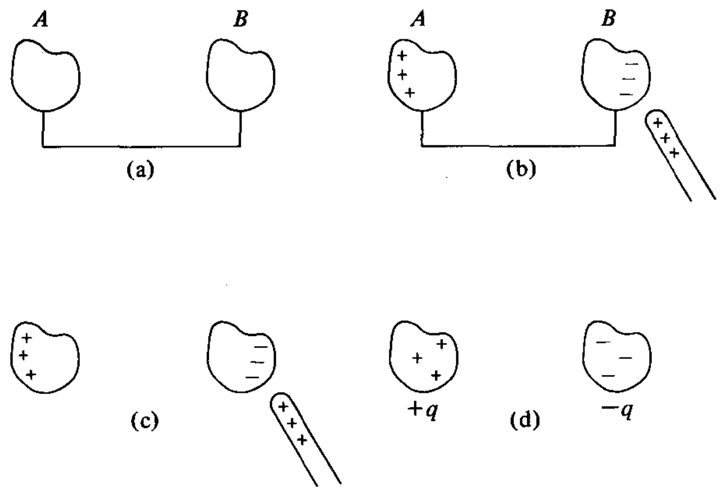


Figure 22-2

charge. As shown in Figure 22-2d, in this way we obtain two conductors carrying equal and opposite charges and thus have a capacitor at a certain charge q.

According to the analysis of Section 21-7, each of the two charged conductors in Figure 22-1 is an equipotential surface and, unless $q = 0$, they are at different potentials. Since the direction of the electric-field lines is from the positively charged conductor A to B, it follows that the potential V_A of A is larger than is the potential V_B of B. The potential V of a capacitor is defined to be the *potential drop* $(V_A - V_B)$ from the positive to the negatively charged conductor. Equivalently, V is the work per unit charge required to take a charged particle from B to A. Note that the potential V of a capacitor is an inherently positive quantity.

The capacitance C of a capacitor is defined by[1]

$$C = \frac{q}{V} \tag{22-1}$$

where V is the potential of the capacitor when it has charge q. Since both V and q are positive quantities it follows that the same must be true for the capacitance C. Note that implicit in this definition is the fact that the capacitance C is a constant independent of V and q. In other words, the ratio q/V for a capacitor is a constant independent of its potential and charge. As will be exemplified for special cases in Section 22-3, C depends only on geometrical factors, such as the shape of the conductors, their relative separation, and so forth.

According to (22-1), the unit of capacitance is the coulomb per volt. A unit equivalent to this is the farad (abbreviated F), which is defined by

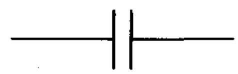
$$1\ \mathrm{F} = 1\ \text{farad} = 1\ \mathrm{C/V}$$

[1]The italic letter C, which stands for capacitance, should not be confused with the roman C, which is the abbreviation for the coulomb, the SI unit of charge. As a general rule, the symbol C for the coulomb will usually be preceded by a number.

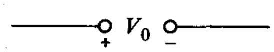
Also in frequent usage is the microfarad (μF), which is defined to be 10^{-6} farad.

Although it is possible to do so, the procedure shown in Figure 22-2 for charging a capacitor is not always the most desirable. A much more convenient and more readily controllable procedure involves the usage of a device, such as a battery, which has two conducting terminals maintained at some potential difference V_0. If these terminals are connected to the plates of a capacitor, charge will flow from one plate to the other until the potential of the capacitor is the same as that between the terminals. If, for example, the potential difference between the terminals is V_0, then when connected to the plates of a capacitor of capacitance C, a charge $+CV_0$ will appear on the plate at the higher potential and a negative charge $-CV_0$ will appear on the other.

It is customary to represent a capacitor schematically by the symbol



where the two vertical lines represent the plates. In a similar way we shall represent any device with two terminals which are maintained at a fixed potential difference V_0 by the symbol



where the + (−) sign represents the terminal at the higher (lower) potential. In terms of these symbols, Figure 22-3 shows a capacitor of capacitance C whose plates are kept at a potential difference V_0. By our convention, the left-hand plate is at the higher potential and has the positive charge CV_0. The other plate has the compensating charge $-CV_0$.

Example 22-1 Suppose that the plates of a 5.0-μF capacitor are connected to a potential difference of 10 volts.

(a) What is the charge on the capacitor?

(b) If the 5.0-μF capacitor is disconnected from the battery and connected to an originally uncharged 3.0-μF capacitor, what is the final charge on each capacitor?

Solution

(a) According to (22-1), the charge is

$$q = CV = 5.0 \times 10^{-6}\,\text{F} \times 10\,\text{V} = 5.0 \times 10^{-5}\,\text{C}$$

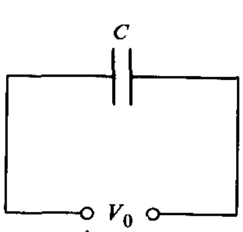


Figure 22-3

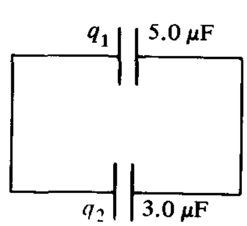


Figure 22-4

(b) If, as shown in Figure 22-4, q_1 represents the final charge on the 5.0-μF capacitor and q_2 that on the other, then $q_1 + q_2 = 5.0 \times 10^{-5}$ coulomb, since the original charge on the 5.0-μF capacitor is 5.0×10^{-5} coulomb. Moreover, since at equilibrium there is no flow of charge, the left-hand plates of the two capacitors must be at the same potential. And for the same reason the two negatively charged plates must be at the same potential. It follows that the potential across each capacitor must be the same, and therefore (since $V = q/C$ for any capacitor)

$$\frac{q_1}{5.0 \times 10^{-6}\ \text{F}} = \frac{q_2}{3.0 \times 10^{-6}\ \text{F}}$$

Solving for q_1 and q_2 we obtain

$$q_1 = 3.1 \times 10^{-5}\ \text{C} \qquad q_2 = 1.9 \times 10^{-5}\ \text{C}$$

22-3 The parallel-plate capacitor

There are a very small number of idealized, but very important, geometric configurations of two conductors for which an explicit formula for the capacitance can be obtained. The purpose of this section is to derive two of these.

Consider first the *parallel-plate* capacitor. This capacitor consists of two very large, parallel, conducting plates, each of area A and separated by certain distance d. If a potential difference V_0 is maintained between the plates, then as shown in Figure 22-5 a certain positive charge per unit area σ will appear on the plate at the higher potential while a compensating negative charge density $-\sigma$ will appear on the other. For regions of space not too close to the edges of the plates, the field lines, as shown in the figure, will originate on the positive charge on the left-hand plate, be perpendicular to the two conductors, and terminate on the negative charge on the other plate. Provided then that the linear dimensions of the plates are very large compared to their separation distance d—a feature we shall always assume

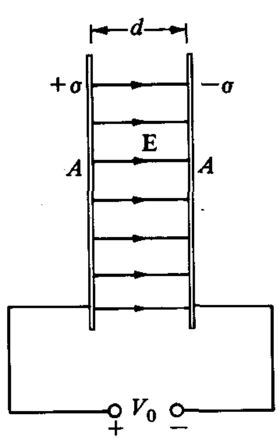


Figure 22-5

to be true—it follows that the electric field will be uniform and perpendicular to both plates and be confined to the region between them. Physically, this means that the nonuniform "fringing fields" that exist at the edges of the plates will be neglected.

Let us then calculate the capacitance C of the parallel-plate capacitor. According to (21-4), the potential V_0 of the capacitor in Figure 22-5 is the negative of the work required to take a unit charge from the left- to the right-hand plate. Since the electric field $\mathbf{E}$ is uniform and perpendicular to the plates it follows that

$$V_0 = Ed$$

Moreover, according to (20-16), E is related to the charge density σ on the positive plate by

$$E = \frac{\sigma}{\epsilon_0}$$

Combining these two relations and noting that the charge q of this capacitor is σA, we find on substitution into (22-1) that

$$C = \frac{q}{V_0} = \frac{\sigma A}{Ed} = \frac{\sigma A}{\sigma d/\epsilon_0}$$

and this leads to the final formula

$$C = \frac{\epsilon_0 A}{d} \tag{22-2}$$

As anticipated above, C is independent of V_0 and q and depends only on the area A of the plates and their separation distance d. Note that the closer together are the plates and the larger is their area, the greater is the capacitance.

The calculation of the capacitance of the spherical and the cylindrical capacitor proceeds, in principle, in the same way. The details will be found in Example 22-4 and in the problems. The results are given by (22-3) and (22-5), respectively.

Example 22-2 Calculate the capacitance of the parallel-plate capacitor whose plates have an area of 15 cm^2 and are separated by a distance of 3.0 mm.

Solution Substituting the given values for A and d into (22-2) and using $\epsilon_0 = 8.9 \times 10^{-12}$ F/m ($\equiv 8.9 \times 10^{-12}$ $C^2/N\text{-}m^2$), we find that

$$C = \frac{\epsilon_0 A}{d} = \frac{(8.9 \times 10^{-12}\ \text{F/m}) \times (15 \times 10^{-4}\ \text{m}^2)}{3.0 \times 10^{-3}\ \text{m}}$$
$$= 4.5 \times 10^{-12}\ \text{F} = 4.5 \times 10^{-6}\ \mu\text{F}$$

Example 22-3 It is desired to place a charge of 1.0 μC on a parallel-plate capacitor by use of a 10-volt source. If the separation distance between the plates is to be 1.0 mm, what must be the area of the plates?

Solution Solving (22-2) for A and substituting for C by use of (22-1), we find that

$$A = \frac{Cd}{\epsilon_0} = \frac{qd}{V\epsilon_0}$$

The substitution of the given values for q, V, d, and ϵ_0 then yields

$$A = \frac{qd}{V\epsilon_0} = \frac{(1.0 \times 10^{-6}\ \mathrm{C}) \times (1.0 \times 10^{-3}\ \mathrm{m})}{(10\ \mathrm{V}) \times (8.9 \times 10^{-12}\ \mathrm{F/m})}$$
$$= 11\ \mathrm{m}^2$$

which is a very large area indeed. We see therefore, that in order to store a charge of the order of 1.0 μC on a capacitor of reasonable size, a substantially larger voltage source must be utilized.

Example 22-4 Calculate the capacitance of a spherical capacitor that consists of two concentric conducting spherical shells of radii a and $b (> a)$, respectively (Figure 22-6).

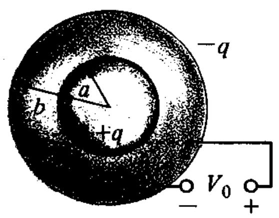


Figure 22-6

Solution As shown in the figure, suppose that there is a charge q on the inner sphere and a compensating charge $-q$ on the outer one. Because of the spherical symmetry, in the region between the conducting shells the potential $V(r)$ is the same as if the charge q were at the center of the inner sphere. Thus for $a \leqslant r \leqslant b$,

$$V(r) = \frac{q}{4\pi\epsilon_0}\frac{1}{r}$$

so that the potential $V[\equiv V(a) - V(b)]$ of the capacitor is

$$V = \frac{q}{4\pi\epsilon_0}\left(\frac{1}{a} - \frac{1}{b}\right)$$

Substitution into (22-1) then yields the desired formula

$$C = 4\pi\epsilon_0 \frac{ab}{(b-a)} \qquad \textbf{(22-3)}$$

For the special case of a very large outer sphere, that is, $b \gg a$, this reduces to

$$C = 4\pi\epsilon_0 a \qquad \left(\frac{b}{a} \to \infty\right) \qquad \textbf{(22-4)}$$

and when speaking of the capacitance of an "isolated" spherical conductor, it is this formula which we have in mind. According to this formula, for example, the capacitance of the earth is about 7×10^{-4} farad.

In the problems it is established that the capacitance of a cylindrical capacitor, consisting of two coaxial cylindrical conductors of radii a and b ($> a$) and of length L, is

$$C = \frac{2\pi\epsilon_0 L}{\ln(b/a)} \tag{22-5}$$

22-4 Capacitors in series and parallel

Two capacitors, C_1 and C_2, can be connected together across a potential difference V_0 in only one of two ways. If they are connected as shown in Figure 22-7a, then they are said to be in *parallel*. If, on the other hand, they are connected as in Figure 22-7b, then they are said to be in *series*. Note that for two capacitors in parallel, the potential across each one is the same, whereas the charges on each are, in general, different. As shown in Figure 22-7a, for the parallel connection the positively charged left-hand plates of *both* C_1 and C_2 are at the same potential as is the positive terminal of the external source, and similarly for the right-hand plates. On the other hand, as shown in Figure 22-7b, for two capacitors in series the positively charged plate of one is connected to and is thus at the same potential as the negatively charged plate of the other. Hence, in general, the potentials of two capacitors in series will be different. However, since the charges on the two plates of a capacitor are equal and opposite, and since the negative charge on the right-hand plate of C_1 can only come from electrons that originate on the left-hand plate of C_2, it follows that the charge on two capacitors in series must invariably be the same.

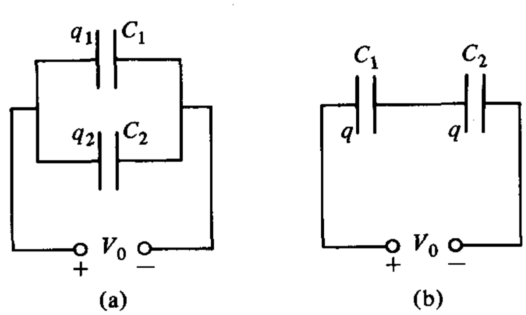


Figure 22-7

We shall now demonstrate that if two capacitors of capacitances C_1 and C_2 are connected in *parallel*, then they are equivalent to a single capacitor of capacitance C, given by

$$C = C_1 + C_2 \qquad \text{(parallel connection)} \tag{22-6}$$

whereas if they are connected in *series*, the corresponding formula is

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} \qquad \text{(series connection)} \tag{22-7}$$

To establish (22-6) let us note that in Figure 22-7a the potential of each capacitor is the same as the potential difference V_0 of the external source. Hence

$$V_0 = \frac{q_1}{C_1} = \frac{q_2}{C_2} \tag{22-8}$$

where q_1 and q_2 are the charges on C_1 and C_2, respectively. Now, by definition, the capacitance C of the capacitor equivalent to these two has the property that when connected to the same source V_0, it will acquire the same *total* charge $(q_1 + q_2)$. Making use of (22-1) we then obtain

$$C = \frac{q_1 + q_2}{V_0} = \frac{C_1 V_0 + C_2 V_0}{V_0} = C_1 + C_2$$

where the second equality follows by use of (22-8). The validity of (22-6) is thereby established.

The derivation of (22-7) follows along similar lines. This time, as noted above and as shown in Figure 22-7b, the charge q is the same on each capacitor. Therefore the potentials of the two capacitors are q/C_1 and q/C_2, respectively. Moreover, from (21-5) and the additivity property of the work carried out in taking a charged particle between two points it follows that the sum of the potentials across the capacitors must be the same as that of the external source, V_0. Hence

$$V_0 = \frac{q}{C_1} + \frac{q}{C_2} = q\left(\frac{1}{C_1} + \frac{1}{C_2}\right)$$

and comparison with (22-1) shows that the capacitance C equivalent to C_1 and C_2 in series is given in (22-7).

Example 22-5 Find the capacitance C of a capacitor equivalent to two capacitors of capacitances $C_1 = 2\ \mu\text{F}$ and $C_2 = 4\mu\text{F}$ if:

(a) They are connected in parallel.

(b) They are connected in series.

Solution

(a) According to (22-6), for the parallel connection we have

$$C = C_1 + C_2 = 2\ \mu\text{F} + 4\ \mu\text{F} = 6\ \mu\text{F}$$

(b) Expressing (22-7) in the form

$$C = \frac{C_1 C_2}{C_1 + C_2}$$

we find for the series connection that

$$C = \frac{(2\ \mu\text{F}) \times (4\ \mu\text{F})}{2\ \mu\text{F} + 4\ \mu\text{F}} = \frac{4}{3}\ \mu\text{F}$$

Example 22-6 Calculate the equivalent capacitance C of the four capacitors connected as in Figure 22-8a. Assume the values $C_1 = C_2 = 2\ \mu\text{F}$ and $C_3 = C_4 = 4\ \mu\text{F}$.

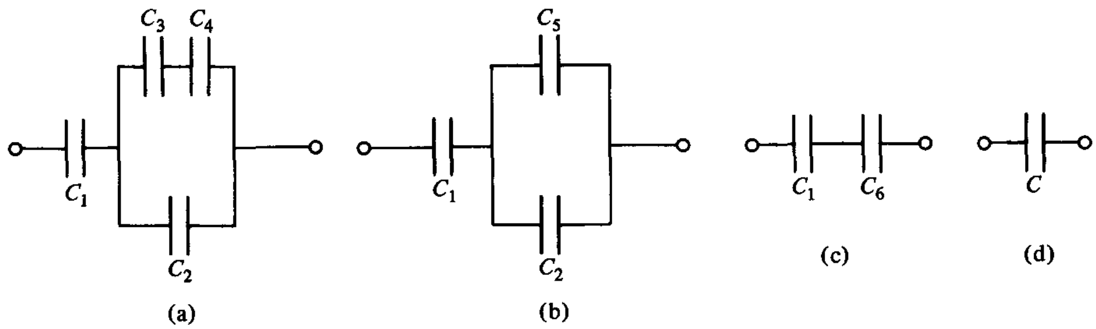


Figure 22-8

Solution Since only parallel and series connections are present, the equivalent capacitance can be found by repeated usage of (22-6) and (22-7).

First, since C_3 and C_4 are in series, they are equivalent to a capacitor of capacitance C_5:

$$C_5 = \frac{C_3C_4}{C_3 + C_4} = \frac{(4\ \mu\text{F}) \times (4\ \mu\text{F})}{4\ \mu\text{F} + 4\ \mu\text{F}} = 2\ \mu\text{F}$$

and the original configuration is reduced to that in Figure 22-8b. Now C_2 and C_5 are in parallel, so by use of (22-6) the configuration in Figure 22-8c results, with C_6 given by

$$C_6 = C_2 + C_5 = 2\ \mu\text{F} + 2\ \mu\text{F} = 4\ \mu\text{F}$$

Finally, combining C_1 and C_6, we obtain the equivalent capacitance C for the original network

$$C = \frac{C_1C_6}{C_1 + C_6} = \frac{(2\ \mu\text{F}) \times (4\ \mu\text{F})}{2\ \mu\text{F} + 4\ \mu\text{F}} = \frac{4}{3}\ \mu\text{F}$$

This example illustrates the method which can be used to reduce certain networks of capacitors to an equivalent single one. Note, however, that *not* all networks of capacitors are reducible by use only of (22-6) and (22-7). The network in Figure 22-9 is an example of this type. Why?

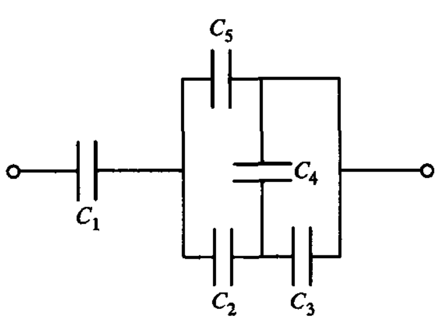


Figure 22-9

22-5 Energy of a charged capacitor

In Chapter 21 the energy stored in the electrostatic field was defined as the work required to assemble any given configuration of charged particles from an initial state of infinite separation. The purpose of this section is to show

that the energy U of a capacitor C with charge Q is

$$U = \frac{1}{2}\frac{Q^2}{C} \tag{22-9}$$

Because of (22-1), equivalent forms for U are

$$U = \frac{QV}{2} = \frac{CV^2}{2} \tag{22-10}$$

where V is the potential of the capacitor when it has charge Q.

According to our definition of electrical energy, the energy U of a capacitor represents the work W required to transfer Q units of charge from one plate of the capacitor to the other. Imagine carrying out this charge transfer in a sequence of steps, each of which involves the transfer of an infinitesimal amount of charge from one plate to the other. If, at some intermediate state, q units of charge have been transferred, then the potential V of the capacitor at this stage is q/C. The work dW required to transport the next infinitesimal charge element dq is then

$$dW = V\,dq = \frac{q}{C}\,dq$$

since the charge dq must be carried through a potential difference $V = q/C$. Integrating both sides of this formula from the initial state, corresponding to $q = 0$, to the final one, $q = Q$, we find that the total work W is

$$W = \int dW = \int_0^Q \frac{q\,dq}{C} = \frac{1}{2C}\,q^2\Big|_0^Q = \frac{Q^2}{2C}$$

Finally, equating this work W to the electrostatic energy U of the capacitor we obtain (22-9).

Example 22-7 A spherical "capacitor" has an inner radius of 0.50 meter and an outer radius of 0.51 meter. How much energy is stored in this capacitor if a potential difference of 100 volts is maintained across its plates?

Solution According to (22-3), the capacitance is

$$C = 4\pi\epsilon_0 \frac{ab}{b-a} = \frac{(0.5\text{ m}) \times (0.51\text{ m})}{(9.0 \times 10^9\text{ N-m}^2/\text{C}^2) \times (0.51\text{ m} - 0.50\text{ m})}$$
$$= 2.8 \times 10^{-3}\ \mu\text{F}$$

Thus it follows by use of (22-10) that the energy is

$$U = \frac{CV^2}{2} = 0.5 \times (2.8 \times 10^{-9}\text{ F}) \times (100\text{ V})^2$$
$$= 1.4 \times 10^{-5}\text{ J}$$

22-6 Force on a capacitor

As an application of the formula for the energy in a capacitor in (22-9) we shall now calculate the strength of the *external* force required to compensate for the attractive electric force that the plates of a charged parallel-plate capacitor exert on each other.

Consider the situation in Figure 22-10, in which the two plates of a capacitor, each of area A, are separated by a distance y and have the respective fixed charges $\pm Q$. We shall assume that the source which originally charged up the capacitor has been disconnected, so that the charge Q on the capacitor is fixed. If f represents the strength of the external force required to keep the plates apart, then the work dW required to increase the separation distance by an amount dy is

$$dW = f\,dy \tag{22-11}$$

On the other hand, according to (22-2) and (22-9), the original energy U of this capacitor is

$$U = \frac{1}{2C} Q^2 = \frac{1}{2} \frac{1}{\epsilon_0 A/y} Q^2 = \frac{1}{2\epsilon_0 A} Q^2 y$$

Therefore, since A and Q^2 are constant, it follows that the change in energy dU as the distance between the plates is increased by dy is

$$dU = \frac{1}{2\epsilon_0 A} Q^2\,dy$$

Finally, equating this to the work dW in (22-11) and canceling out the factor dy, we obtain

$$f = \frac{Q^2}{2\epsilon_0 A} \tag{22-12}$$

The fact that this external force is positive reflects the fact that it must be directed, as shown in the figure, in a way to keep the plates from coming together.

Even though the formula for f in (22-12) has been derived on the

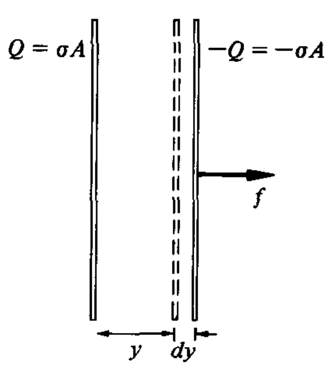


Figure 22-10

assumption that the charge on the capacitor is fixed—that is, that Q does not change as the separation distance between the plates is increased—it is also valid if a fixed potential V is maintained across the plates. This second possibility is more complex since the external source is also able to carry out work and thus it is not possible in this case simply to equate the external work $f\,dy$ to the energy change dU of the electrostatic field. However, making use of the fact that according to (21-5) the work carried out in transporting a charge dq is $V\,dq$, it is shown in the problems that (22-12) is valid in this more complex case nevertheless.

Example 22-8 Suppose a 100-volt potential difference is maintained between the plates of a 1.0-μF parallel-plate capacitor. What external force is required to keep the plates at the fixed separation distance of 10^{-3} meter?

Solution Making use of (22-12), (22-1), and (22-2) we have

$$f = \frac{Q^2}{2\epsilon_0 A} = \frac{Q^2}{2Cd} = \frac{1}{2d}CV^2$$

and the substitution of the given parameter values leads to

$$\begin{aligned} f = \frac{1}{2d}CV^2 &= \frac{1}{2.0 \times 10^{-3}\,\mathrm{m}} \times (1.0 \times 10^{-6}\,\mathrm{F}) \times (100\,\mathrm{V})^2 \\ &= 5.0\,\mathrm{N} \end{aligned}$$

22-7 Dielectrics—qualitative discussion

Having established some of the important properties of capacitors, in the remainder of this chapter we shall make use of this knowledge to study dielectrics.

By contrast to a solid conductor whose electrons are relatively free to wander through a fixed lattice of positive ions, the electrons in an insulator, as noted previously, are *not* able to move freely. Rather, we find that each such electron is strongly bound to its parent ion and is thereby permanently attached to it. In the absence of an external electric field, then, each electron in an insulator may be thought of as orbiting about its parent ion and in such a way that, overall, each atom is electrically neutral.

A typical value for the velocity of an electron as it moves about its parent ion is of the order of 10^6 m/s. Since the atomic radius is of the order of 10^{-10} to 10^{-9} meter, this means that an electron will make about 10^{15} to 10^{16} traversals about its associated ion in 1 second. On a macroscopic scale, only time averages of such high-speed microscopic motions can be observed. Hence the physical picture that emerges—which, incidentally, is very similar to that predicted by quantum mechanics—is that each atom of a dielectric consists of a positive core (the ion) surrounded by a "smear" of negative charge. Generally speaking, this smear of negative charge will be distributed

symmetrically and its center of gravity will coincide with that of the positively charged ion. An application of Gauss' law then shows that the electric field outside of such an atom vanishes, and that the atom must therefore be electrically neutral. Figure 22-11 depicts such a macroscopic view of a hydrogen atom. Note that the center of gravity of the negative-charged cloud is located at the position of the nucleus.

Consider now what happens to an atom of a dielectric when it is in the presence of an external electric field $\mathbf{E}_0$. The electron will again travel about the ion, but this time its orbit will not have the spherically symmetric form in Figure 22-11. Instead, because of the existence of the external field, the negative cloud will become distorted and will assume a shape such as that in Figure 22-12. The reason for this distortion is as follows. Since the electron has a negative charge, the downward electric force it experiences when it is at, say, point A is larger than is the upward force on it when it is at point B. For at A the force due to the external electric field $\mathbf{E}_0$ and that due to the ion are in the same direction, whereas at B these forces are oriented in opposite directions. It follows that a time-averaged view of this electronic motion must have a structure similar to that in the figure.

As implied in Figure 22-12, in the presence of an electric field the center of gravity of the negative electron cloud does *not* coincide with that of the ion. This means that there is an effective charge separation between the electron and the ion, as a result of which the atom acquires a certain electric dipole moment $\mathbf{p}$. We shall use the term *polarization* to characterize this property of an atom's acquiring an electric dipole moment when placed into an electric field. According to Figure 22-12, the induced dipole moment $\mathbf{p}$, which is conventionally defined to be parallel to a vector from the negative to the positive charge, is parallel to the direction of the external field $\mathbf{E}_0$. Moreover, experiment shows further that for electric fields of moderate strength this induced dipole moment $\mathbf{p}$ is proportional to the strength of the electric field. Thus

$$\mathbf{p} = \alpha \mathbf{E}_0 \tag{22-13}$$

with α a proportionality constant known as the *atomic polarizability*. Table 22-1 lists the values[2] of α for typical atoms. Note that for the tightly bound

Table 22-1 Atomic polarizabilities in units of 10^{-40} (C-m^2)/V

Element	H	He	Li	C	Ne	Na	Ar	K
α	0.73	0.23	13	1.7	0.44	30	1.8	38

[2]Strictly speaking, the table lists the *static* values for α. If the external field varies in time, (22-13) is still valid, but with the value of α determined by the details of this variation.

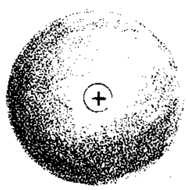


Figure 22-11

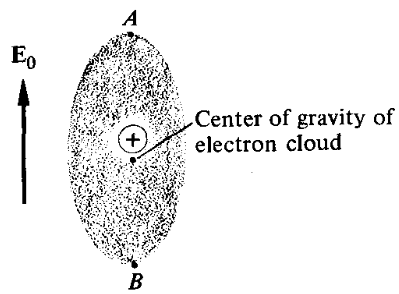


Figure 22-12

electrons in an inert gas, such as He, Ne, or Ar, the polarizability is very small compared with the values of α for the alkalis, such as Li, Na, and K, whose electrons are more loosely bound.

Consider, for example, the case of a hydrogen atom in an external field of strength 10^4 V/m. Using the value for α listed in the table, we find that the induced dipole moment p has the value

$$p = \alpha E = \left(7.3 \times 10^{-41} \frac{\text{C-m}^2}{\text{V}}\right) \times \left(10^4 \frac{\text{V}}{\text{m}}\right) = 7.3 \times 10^{-37} \text{ C-m}$$

Since the electric charge q of this dipole has the value 1.6×10^{-19} coulomb, the effective separation distance l between the proton and the center of gravity of the electron cloud is

$$l = \frac{p}{q} = \frac{7.3 \times 10^{-37} \text{ C-m}}{1.6 \times 10^{-19} \text{ C}} = 4.6 \times 10^{-18} \text{ m}$$

This corresponds to a very slight distortion of the originally spherically symmetric electronic orbit and is of the order of 10^{-7} that of the atomic radius.

Having made plausible the fact that in the presence of an electric field the individual atoms become polarized, let us consider what happens when an insulator is placed into an electric field. As for the isolated atom above, each atom in the dielectric will become polarized and set up its own dipole field. However, the individual dipoles in the dielectric will not, in general, line up along the direction of the external field! For the electric force acting on a given atom is due not only to the external field but to the dipole fields produced by its nearby neighbors as well. In general, these two fields are not parallel. In addition, there are effects associated with the thermal vibrations of the atoms that also tend to counteract the tendency for the dipoles to line up along $\mathbf{E}_0$. However, for isotropic and homogeneous dielectrics—the only substances with which we shall be concerned—it is valid to think of the dipoles in the dielectric as being lined up with the external field, and we shall do so in all of the following.

22-8 The dielectric constant

The purpose of this section is to define a certain parameter κ, called the *dielectric constant*, in terms of which the macroscopic, electric properties of isotropic, homogeneous insulators may be described. The relation between κ and the dipole moments **p** of the constituent atoms and molecules will be discussed in the following sections.

Consider a parallel-plate capacitor of area A and separation distance between the plates d. According to (22-2), its capacitance, for which we now use the symbol C_0, is

$$C_0 = \frac{\epsilon_0 A}{d} \tag{22-14}$$

Imagine now carrying out various measurements on this capacitor to determine the effect of placing between its plates a slab of dielectric of area A and of thickness d. First, as shown in Figure 22-13, let us insert the slab

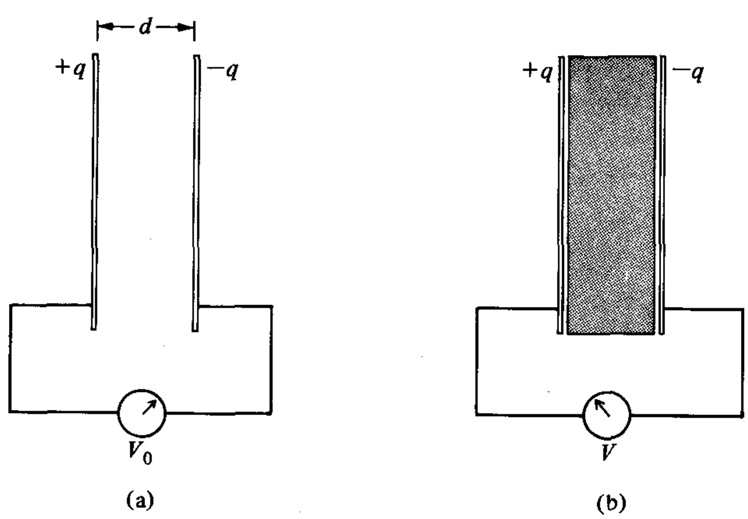


Figure 22-13

after the capacitor has first been given a charge q and has then been disconnected from the line. A measurement of the potential difference across the plates shows that the potential V of the capacitor with the dielectric in is less than the original value $V_0 = q/C_0$. Second, as shown in Figure 22-14, let us insert the dielectric slab while a fixed potential difference V is maintained across the plates. This time, the final charge q on the capacitor is found to exceed the original charge $q_0 = C_0 V$. Moreover, comparison of the results of these two experiments—one with a fixed charge on the capacitor and the second with a fixed potential across its plates—shows that, as the dielectric is inserted, the factor by which the potential decreases in the first case is precisely the same as the factor by which the charge increases in the second. In other words, the ratio V_0/V of the

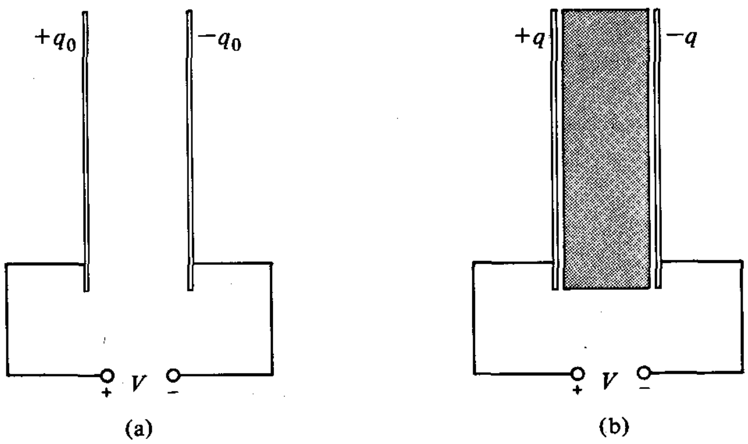


Figure 22-14

potential difference measurements in Figure 22-13 has precisely the same value as does the ratio q/q_0 of the charges in Figure 22-14.

To interpret these results, it is useful to turn to the defining relation for capacitance in (22-1). Evidently the insertion of the dielectric slab between the plates of the capacitor has increased its capacitance by the equivalent factors V_0/V or q/q_0 in the two cases above. Thus, we shall write

$$C = \kappa C_0 \tag{22-15}$$

where $C_0 = \epsilon_0 A/d$ is the capacitance of the capacitor in the absence of the dielectric slab and C is the corresponding capacitance after it has been inserted. The coefficient κ is called the dielectric constant for the given material. Note that κ is a dimensionless number and satisfies the inequality $\kappa \geqslant 1$.

If the above experiments are repeated with the same (homogeneous and isotropic) dielectric material but with capacitors of various sizes and shapes, (22-15) is found to be valid for these cases also. This implies that the dielectric constant κ, as defined above, *is a property of the given material only.* In particular, it does not depend on the charge or the potential of the capacitor, nor does it depend on its geometrical structure.

Table 22-2 lists some experimental values for the dielectric constants of a number of materials. Note that $\kappa > 1$ for all of them.

Table 22-2 Dielectric constants at 20°C

Substance	*Air*	*Beeswax*	*Paraffin*	*Pyrex*	*Water*
κ	1.0006	2.9	2.3	4.0	80

Example 22-9 Suppose that a 10-volt source is connected across a 2.0-μF parallel-plate capacitor.

(a) What is the charge on the capacitor?

(b) Suppose that the source is disconnected and a dielectric slab is introduced between the plates. If in this process the potential across the capacitor drops to 6.0 volts, what is the dielectric constant κ?

(c) What is the capacitance of the capacitor with the dielectric slab in?

Solution

(a) According to (22-1), the original charge q is

$$q = C_0 V_0 = (2.0 \times 10^{-6}\ \text{F}) \times (10\ \text{V}) = 2.0 \times 10^{-5}\ \text{C}$$

(b) Since the charge is fixed, it follows from (22-1) that the product CV is independent of the presence or absence of the slab. According to (22-15), the capacitance increases by the factor κ and thus the potential across the capacitor must decrease by the same factor. Using the given parameter values, we thus obtain

$$\kappa = \frac{10\ \text{V}}{6\ \text{V}} = 1.7$$

(c) Substituting this value for κ into (22-15), we find that

$$C = \kappa C_0 = 1.7 \times 2.0\ \mu\text{F} = 3.4\mu\text{F}$$

22-9 Polarization charge

In order to relate the dielectric constant of an insulator to an induced charge, consider again the physical situation in Figure 22-13 of a capacitor with a fixed charge q. According to (22-15), the insertion of the dielectric slab between the plates increases its capacitance by the factor κ or, equivalently, decreases the potential difference between the plates by this factor. But for a uniform electric field **E**, the potential difference between any two equipotential planes a distance d apart is Ed. Hence, from the fact that the potential difference between the plates in Figure 22-13 decreases by the factor κ when the slab is inserted, it follows that the electric field in this region must also decrease by this factor. Thus if E_0 represents the magnitude of the uniform electric field between the plates *before* the dielectric slab is inserted and E_i is the corresponding value afterward, then

$$E_i = \frac{E_0}{\kappa} \tag{22-16}$$

Moreover, if the steps leading to (22-2) are repeated but with E replaced by this value E_i, then since $E_0 = \sigma/\epsilon_0 = q/\epsilon_0 A$, the defining relation in (22-15) results, as it should. We shall make extensive use of (22-16) in all of the following.

Further confirmation of the validity of (22-16) may be obtained in the following way. In Example 22-10, it is established that the capacitance C of the capacitor in Figure 22-15, whose plates have an area A and are separated by a distance d, and which contains a dielectric slab of constant κ and

thickness t ($< d$), is

$$C = \frac{\epsilon_0 A}{(d-t)+t/\kappa} \tag{22-17}$$

In deriving this relation we shall find it necessary to make essential use of (22-16). Hence the experimental verification in the laboratory of (22-17) for all thicknesses t ($< d$) is in effect a confirmation of the validity of (22-16).

The physical significance of the internal field E_i is illustrated in Figure 22-16. As shown in Example 22-12, on the face of the dielectric slab nearest the positive plate of the capacitor there is induced a negative polarization charge density, $-\sigma_p$, where

$$\sigma_p = \epsilon_0 \frac{\kappa - 1}{\kappa} E_0 \tag{22-18}$$

and a compensating charge density $+\sigma_p$ is induced on the opposite face. For the isotropic and homogeneous dielectrics of interest to us, no polarization charge comes into existence in the interior of the dielectric. As shown in the figure, some of the field lines that originate on the positive plate of the capacitor terminate on the polarization charge on the left face of the dielectric while others continue on through to the charge on the negative plate. The field lines that originate on the positive charge on the right face of the dielectric terminate on the negative plate of the capacitor. Both of these features are consistent with (22-16) and its implication that $E_i < E_0$.

Example 22-10 Derive (22-17).

Solution Suppose that the capacitor in Figure 22-15 has a charge q. According to (20-16), the electric field E_0 in the region between the plates but *outside* of the dielectric is

$$E_0 = \frac{\sigma}{\epsilon_0} = \frac{q}{\epsilon_0 A}$$

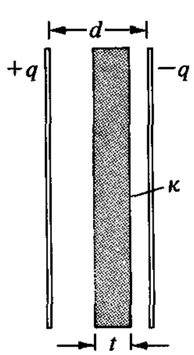


Figure 22-15

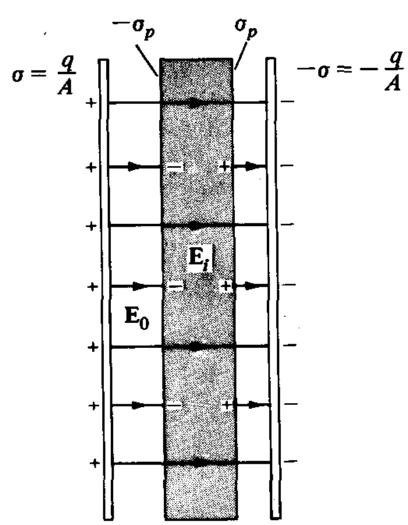


Figure 22-16

which is directed to the right in the figure if the plate on the left has the positive charge. The field $\mathbf{E}_i$ inside the dielectric lies along the same direction and, according to (22-16), has the magnitude

$$E_i = \frac{E_0}{\kappa} = \frac{q}{\epsilon_0 \kappa A}$$

Making use of these formulas and the result in Example 21-13, we find that the potential V across the capacitor is

$$V = E_0(d-t) + E_i t = E_0\left[(d-t) + \frac{t}{\kappa}\right]$$

$$= \frac{q}{\epsilon_0 A}\left[(d-t) + \frac{t}{\kappa}\right]$$

Finally, solving for the ratio q/V and making use of the definition in (22-1), we obtain (22-17).

Example 22-11 For the dielectric slab in Figure 22-16, suppose that $\kappa = 2$ and that the capacitor is charged so that $E_0 = 100$ V/m. Calculate the field inside the dielectric and the surface polarization charge on the dielectric.

Solution Substituting the given data into (22-16), we find that

$$E_i = \frac{E_0}{\kappa} = \frac{100\ \text{V/m}}{2} = 50\ \text{V/m}$$

The surface polarization charge $\sigma_p{}^l$ on the left side of the dielectric is negative and, according to (22-18), has the value

$$\sigma_p{}^l = -\sigma_p = -\frac{\epsilon_0(\kappa - 1)}{\kappa} E_0$$

$$= -\left(8.9 \times 10^{-12} \frac{\text{C}^2}{\text{N-m}^2}\right) \times \left(\frac{2-1}{2}\right) \times \left(\frac{100\ \text{V}}{\text{m}}\right)$$

$$= -4.5 \times 10^{-10}\ \text{C/m}^2$$

A similar calculation shows that an equal charge density, but of the opposite sign, is induced on the right-hand face of the slab.

Example 22-12 Derive (22-18).

Solution Consider, in Figure 22-17, one face of a homogeneous and isotropic dielectric of constant κ in a uniform external field $\mathbf{E}_0$ perpendicular to this face.

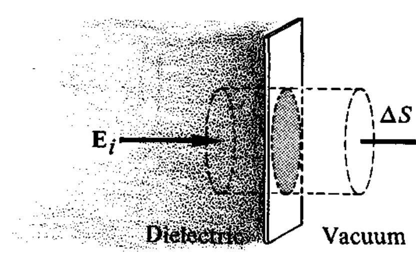


Figure 22-17

Inside the dielectric there will be an electric field $\mathbf{E}_i$ also perpendicular to the surface and with magnitude given in (22-16). To calculate the polarization charge density σ_p induced on the dielectric surface, let us construct a Gaussian surface in the shape of a cylinder of cross-sectional area ΔS with its axis parallel to $\mathbf{E}_0$. Reference to Figure 22-17 shows that since one end face of the cylinder is inside the dielectric and the other outside, the total flux out of the cylinder is $(E_0\,\Delta S - E_i\,\Delta S)$. Therefore, since the total charge inside is $\sigma_p\,\Delta S$, it follows from Gauss' law that

$$(E_0 - E_i)\,\Delta S = \frac{1}{\epsilon_0}\,\sigma_p\,\Delta S$$

The result in (22-18) then follows by canceling the factor ΔS and substituting for E_i by use of (22-16).

22-10 Energy considerations

The purpose of this section is to extend the energy considerations of Section 22-5 to the case of a parallel-plate capacitor that contains a dielectric slab.

Consider first the case of a capacitor without a dielectric but with a fixed charge Q. According to (22-9), its energy U_0 is

$$U_0 = \frac{1}{2C_0}\,Q^2$$

with C_0 given in (22-14). If a dielectric slab of constant κ is introduced between the plates, then, according to (22-15), the capacitance is increased by the factor κ. Since the charge on the capacitor is assumed to be fixed and is thus unaltered in this process, it follows that the final energy U is

$$U = \frac{1}{2C}\,Q^2 = \frac{1}{2\kappa C_0}\,Q^2 = \frac{1}{\kappa}\,U_0 \tag{22-19}$$

with U_0 the energy of the capacitor without the dielectric. Since the work W required to insert the dielectric between the plates is equal to the change in energy $(U - U_0)$, it follows by use of (22-19) that

$$W = U - U_0 = \frac{1}{\kappa}\,U_0 - U_0 = -U_0\left(\frac{\kappa - 1}{\kappa}\right) \tag{22-20}$$

The fact that this work is negative, since $\kappa > 1$, means that the electrostatic field carries out positive work on the dielectric and thus pulls it into the region between the plates. The physical mechanism that underlies this effect is the fringing field at the edge of the capacitor. Figure 22-18 shows schematically some of the field lines at the edges of a capacitor and thus illustrates how it is possible for the dielectric to be pulled into the region between the plates.

Let us now reexamine this problem, but suppose this time that the potential across the plates is kept at a fixed value V. In the absence of a dielectric, the energy U_0 stored in the capacitor is $C_0V^2/2$, with C_0 defined in

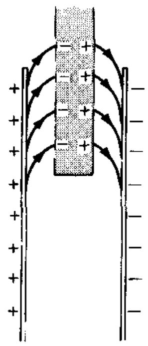


Figure 22-18

(22-14). With the insertion of the dielectric, the capacitance C is increased by the factor κ, and thus the final energy U is

$$U = \frac{CV^2}{2} = \frac{\kappa C_0 V^2}{2} = \kappa U_0$$

This corresponds to an energy change

$$U - U_0 = \kappa U_0 - U_0 = (\kappa - 1)U_0 \qquad \textbf{(22-21)}$$

which, by contrast to the fixed-charge case in (22-20), is positive. However, this increase in energy does *not* necessarily mean that the external agent must carry out positive work to insert the dielectric. For the external source that keeps the plates at the potential V will also, in general, carry out work while the dielectric is being inserted.

To calculate the work W_s carried out by the source, let us note that the charge Q of the capacitor after the dielectric has been inserted is

$$Q = CV = \kappa C_0 V = \kappa Q_0 \qquad \textbf{(22-22)}$$

where Q_0 is the charge without the dielectric. This means that a net charge $(Q - Q_0) = (\kappa - 1)Q_0$ flows as a result of the insertion of the dielectric slab. Hence the work W_s carried out by the source at the fixed potential V is

$$\begin{aligned} W_s &= V(Q - Q_0) = VQ_0(\kappa - 1) \\ &= C_0 V^2(\kappa - 1) \\ &= 2U_0(\kappa - 1) \end{aligned} \qquad \textbf{(22-23)}$$

where the last equality follows since the original energy U_0 is $C_0V^2/2$. Now if W is the work required of an external agent to insert the dielectric into the capacitor, then according to the energy-conservation principle

$$W_s + W = U - U_0$$

which in words states that the change in electrostatic energy $(U - U_0)$ must come from the work produced by *all* external sources. Substituting (22-21) and (22-23) into this formula we find that the external work W is

$$W = U - U_0 - W_s = U_0(\kappa - 1) - 2U_0(\kappa - 1) = -U_0(\kappa - 1) \qquad \textbf{(22-24)}$$

The fact that this is negative implies that the dielectric will be pulled into the region between the plates, as in the other case.

To summarize, then:

If a dielectric of constant κ is inserted between the plates of a capacitor which is kept at a fixed potential V, then the battery carries out a positive amount of work $2(\kappa - 1)U_0$, where U_0 is the original energy. Half of this energy shows up as an increase of the electrostatic field energy, and the other half is expended in the form of work required to draw the dielectric into the field.

Example 22-13 A 3.0-μF parallel-plate capacitor is connected across a 100-volt line. If a dielectric of constant $\kappa = 2.2$ is inserted between the plates, calculate:

(a) The original electrostatic energy.
(b) The work carried out by the battery as the dielectric is inserted.
(c) The increase in field energy.
(d) The final charge on the capacitor.

Solution

(a) The original energy U_0 is

$$U_0 = \frac{1}{2} C_0 V^2 = \frac{1}{2} \times (3.0 \times 10^{-6}\ \text{F}) \times (100\ \text{V})^2$$
$$= 1.5 \times 10^{-2}\ \text{J}$$

(b) The work W_s carried out by the source is, according to (22-23),

$$W_s = 2U_0(\kappa - 1) = 2 \times 1.5 \times 10^{-2}\ \text{J} \times (2.2 - 1) = 3.6 \times 10^{-2}\ \text{J}$$

(c) Of the 3.6×10^{-2} joule expended by the source, one half, or 1.8×10^{-2} joule, goes to increase the field energy from its original value of 1.5×10^{-2} joule to its final value of 3.3×10^{-2} joule.

(d) The original charge on the capacitor Q_0 is

$$Q_0 = C_0 V = (3.0 \times 10^{-6}\ \text{F}) \times 100\ \text{V} = 3.0 \times 10^{-4}\ \text{C}$$

and thus, by use of (22-22), we obtain, for the final charge,

$$Q = \kappa Q_0 = 2.2 \times 3.0 \times 10^{-4}\ \text{C} = 6.6 \times 10^{-4}\ \text{C}$$

†22-11 The displacement vector

In the above discussion of dielectrics we considered only those substances which could be characterized as being homogeneous and isotropic dielectrics, and treated only those cases for which the external field was uniform and directed perpendicularly to the dielectric surface. The purpose of this

section is to describe briefly how to deal with situations for which these conditions are not satisfied.

If a dielectric is introduced into an electric field, then, in general, its constituent molecules become polarized and tend to line up with the resultant field. It is convenient to describe this electrical behavior of the dielectric in terms of a certain quantity known as the *dipole moment per unit volume* **P**. This quantity **P** represents the macroscopic average of the individual molecular dipole moments **p**, and is defined so that $\mathbf{P}\Delta V$ represents the dipole moment associated with any given small volume element ΔV. In general, **P** varies from point to point in space and is nonvanishing only inside and on the surface of dielectric materials.

On physical grounds we expect that if a dielectric is inserted into an electric field, a certain polarization charge will be induced within, and on the surface of, the sample. Let $q_p(S)$ represent this polarization charge contained in an arbitrary, but fixed, closed surface S (Figure 22-19). The relation between $q_p(S)$ and the dipole moment per unit volume **P** is

$$q_p(S) = -\oint_S \mathbf{P} \cdot d\mathbf{S} \tag{22-25}$$

where the integral is over the closed surface S. The validity of this relation for the special case for which **P** is constant inside the dielectric will be confirmed in the problems.

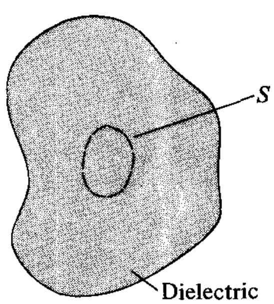


Figure 22-19

Consider now a dielectric characterized by a dipole moment per unit volume **P** and let **E** represent the electric field produced by the associated polarization charge plus any other charge that may be present. We define the *displacement vector* **D** associated with this system by

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{22-26}$$

so that outside of the dielectric, where $\mathbf{P} = 0$, the displacement vector **D** is—up to a factor ϵ_0—precisely the same as is the electric field **E**. However, inside the dielectric where, in general, $\mathbf{P} \neq 0$, the relation between **D** and **E** is given by the more complex form in (22-26). We shall now give a physical meaning to **D** by showing that it represents that part of the electric field produced by all electric charges except those induced in the dielectric.

To this end, consider an arbitrary closed surface S in a region of space that may contain both ordinary charge as well as polarization charge which it induces in dielectric materials. Let $q(S)$ represent the ordinary or "true" charge inside S and $q_p(S)$ the polarization charge contained within this same surface. The total charge contained within S is the sum of these; that is, $[q(S)+q_p(S)]$. Applying Gauss' law to the surface S, we have

$$\oint_S \mathbf{E}\cdot d\mathbf{S}=\frac{1}{\epsilon_0}[q(S)+q_p(S)] \tag{22-27}$$

where $\mathbf{E}$ is the total electric field and is produced by all charges. Let us now take the dot product of both sides of (22-26) with an infinitesimal area element $d\mathbf{S}$ and integrate over the closed surface S. The result is

$$\oint \mathbf{D}\cdot d\mathbf{S}=\epsilon_0\oint \mathbf{E}\cdot d\mathbf{S}+\oint \mathbf{P}\cdot d\mathbf{S}$$

and, substituting for the surface integrals on the right-hand side by use of (22-25) and (22-27), we find that

$$\oint_S \mathbf{D}\cdot d\mathbf{S}=\epsilon_0\left\{\frac{1}{\epsilon_0}[q(S)+q_p(S)]\right\}-q_p(S)$$

which simplifies to

$$\oint_S \mathbf{D}\cdot d\mathbf{S}=q(S) \tag{22-28}$$

Comparison with Gauss' law shows then that up to a factor ϵ_0 the *flux of the displacement vector* **D** *out of the closed surface S is equal to the "true" charge inside S.* This leads to the physical interpretation for **D** as that part of the electric field produced by the "true" charge, that is, the nonpolarization charge $q(S)$. Since in general we have much more knowledge about $q(S)$ than we do about $q_p(S)$, the calculation of **D** is often much simpler than is that of **E**. This is one of the main reasons for introducing this notion of the displacement vector.

For many cases of physical interest the induced dipole moment per unit volume **P** is proportional to the electric field **E**. Materials for which this relation is valid are known as *linear materials*. For a linear material, then,

$$\mathbf{P}=\epsilon_0\chi_e\mathbf{E} \tag{22-29}$$

where the dimensionless coefficient of proportionality χ_e is known as the *electric susceptibility*. Outside of the dielectric χ_e vanishes, of course. Experiment shows that the quantity χ_e is related to the dielectric constant of the material by

$$\chi_e=\kappa-1 \tag{22-30}$$

Indeed, this relation is often taken as the definition of the electric susceptibility.

Additional confirmation of the validity (22-30) may be obtained in the

following way. Take the dot product of (22-29) with $d\mathbf{S}$ and integrate over the closed dashed surface in Figure 22-17. According to (22-25), we obtain on the left-hand side $(-\sigma_p\,\Delta S)$. On the right-hand side, since $\chi_e = 0$ outside of the dielectric we obtain, similarly, $-\epsilon_0\chi_e E_i\,\Delta S$. Equating these two, we obtain $\sigma_p = \epsilon_0\chi_e E_i$, and this leads directly to (22-30) after comparison with (22-16) and (22-18).

For linear materials, for which (22-29) is applicable, there is a very simple relation between **D** and **E**. Eliminating **P** and χ_e between (22-26), (22-29), and (22-30), we obtain

$$\mathbf{D} = \kappa\epsilon_0\mathbf{E} \tag{22-31}$$

22-12 Summary of important formulas

A capacitor is a device consisting of two isolated conductors having equal and opposite charges q (>0) and $-q$, respectively. If V is the potential of the positive plate relative to the negative one, then the capacitance C is an inherently positive quantity, defined by

$$C = \frac{q}{V} \tag{22-1}$$

For a parallel-plate capacitor of area A and plate separation distance d, C is

$$C = \frac{\epsilon_0 A}{d} \tag{22-2}$$

The electrostatic energy U stored in a capacitor of capacitance C is

$$U = \frac{Q^2}{2C} \tag{22-9}$$

where Q is the charge on the capacitor. By use of (22-1), various equivalent forms for U can be obtained.

If a dielectric slab of constant κ is inserted between the plates of a capacitor so that it occupies all available space, then its capacitance C is

$$C = \kappa C_0 \tag{22-15}$$

QUESTIONS

1. Define or describe briefly the following: **(a)** parallel-plate capacitor; **(b)** spherical capacitor; **(c)** dielectric; **(d)** dielectric constant; and **(e)** fringing field.
2. According to (22-2), the capacitance of a parallel-plate capacitor can be made arbitrarily large by decreasing the separation distance d between the plates. What practical considerations set an upper limit to the capacitance of such a capacitor?
3. Why are the two conductors A and B in Figure 22-2b at the same potential even though they have equal and opposite charges? Why are they still

at the same potential in Figure 22-2c?

4. If C_0 is the capacitance of the two conductors in Figure 22-1 in a vacuum, present a qualitative argument showing that if the intervening space is filled with a dielectric of constant κ, then the capacitance is increased to the value κC_0.
5. Is it necessary that the inner conducting sphere in Figure 22-6 be hollow? What would the capacitance be if it were solid?
6. Why is it necessary to supply an external force to keep the plates of a parallel-plate capacitor from coming into contact?
7. Show explicitly that the unit of ϵ_0 is the farad per meter (F/m).
8. If two charged capacitors are connected together so that the positive plate of one is in contact with the positive one of the other, and similarly for the negative plates, are the capacitors in series or in parallel?
9. Repeat Question 8, but suppose this time that the positive plate of each one is connected to the negative plate of the other.
10. Consider a parallel-plate capacitor of area A. Is it legitimate to think of it as consisting of two parallel-plate capacitors each of area $A/2$ and connected together? Are they connected in series or in parallel?
11. A flat, metal plate is inserted between the plates of a parallel-plate capacitor, as in Figure 22-20. Is the capacitance increased or decreased thereby? What is the sign of the work that must be carried out to insert the metal plate?

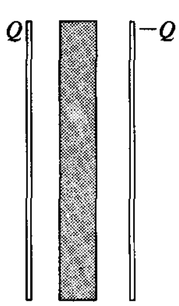


Figure 22-20

12. Would the capacitance of the capacitor in Figure 22-20 increase or decrease if the central conductor were connected to one of the plates by a conducting wire?
13. Describe in what sense the configuration in Figure 22-20 may be thought of as two capacitors connected in series.
14. List some practical difficulties associated with the calculation of the capacitance of two spherical conductors whose centers are separated by a distance greater than the sum of their radii.
15. Explain why the capacitance C of two conductors separated by a distance L, which is very large compared to their linear dimensions, is

$$C = 4\pi\epsilon_0 L$$

(*Hint*: What would be the potential difference between these conductors if they had charges q and $-q$?)

16. Two identical capacitors are separately charged up to the same potential, and then disconnected from the voltage source and reconnected together so they discharge. What happens to the energy stored in the capacitors? Explain.
17. A parallel-plate capacitor is charged to a potential V and lies on a horizontal surface. Neglecting friction, describe qualitatively the subsequent motion of a dielectric slab that is introduced at one end of the capacitor.
18. A parallel-plate capacitor has a fixed charge. If a dielectric slab is introduced between the plates, describe what happens to the following: (a) the potential across the plates; (b) the capacitance; (c) the electric field

strength; (**d**) the energy; and (**e**) the charge on the plates.

19. Repeat Question 18, but this time assume that a fixed potential difference is maintained across the plates.

20. Explain why in the interior of a dielectric slab in which there exists a uniform electric field there can be no electric charge.

21. Two capacitors C_1 and C_2 have the respective charges q_1 and q_2 and are connected together in parallel. If a dielectric slab is inserted into C_2, what happens to (**a**) q_1 and q_2? (**b**) C_1 and C_2? (**c**) The potential across the capacitors?

22. Show that the total energy stored in two capacitors connected in series across a certain source is the same as the energy stored in the equivalent capacitor when connected across the same source.

23. Repeat Question 22, but this time assume that the two capacitors are connected in parallel across the source.

24. Show, by use of (22-18), that under some circumstances a conductor may be thought of as a dielectric of infinite dielectric constant.

PROBLEMS

1. If a 6.0-volt source is connected across a 2.0-μF capacitor, how much charge will be transferred between the conductors?

2. If a charge of 0.3 μC appears on the positive plate of a capacitor when a 100-volt source is connected across the plates, what is its capacitance?

3. The area of each plate of a parallel-plate capacitor is 20 cm^2 and the plates are separated by a distance of 1.5 mm. (**a**) What is the capacitance? (**b**) If a 20-volt potential is connected across the capacitor, what *charge density* appears on each plate?

4. The distance between the plates of a parallel-plate capacitor is doubled. (**a**) By what factor must the area be increased to keep the capacitance fixed? (**b**) What must be the dielectric constant of an insulator that if placed between the plates keeps the capacitance unaltered?

5. A spherical capacitor consists of two concentric spheres of radii 10 cm and 9.9 cm, respectively.

(**a**) What is the capacitance?

(**b**) If this capacitor is connected to a 10-volt source, what is the total charge on the positive sphere?

(**c**) Assuming that the larger sphere is at the higher potential, calculate the surface charge density on the two spheres. Explain why these two densities are *not* equal and opposite.

6. A parallel-plate capacitor of area 12 cm^2 and separation distance 2.0 mm has a charge of 0.1 μC. (**a**) What is the potential difference between the plates? (**b**) What is the electric field strength?

7. Suppose a dielectric slab of constant $\kappa = 1.5$ is inserted between the plates of the capacitor in Problem 6. (**a**) What is the potential of the capacitor now? (**b**) What is the electric field strength in the dielectric?

8. Calculate the electrostatic energy stored in the capacitor in Problem 1 when it is fully charged? Explain why this is *not* the same as the work carried out by the 6.0-volt source that was used to charge it up.

9. Assuming the sun to be a conducting sphere of diameter 1.4×10^6 km, calculate its capacitance.
10. What must be the capacitance of a capacitor that will store 10 joules of energy if connected to a 100-volt potential difference? What would the area of its plates have to be if it were a parallel-plate capacitor with a separation distance of 5.0×10^{-4} meter?
11. What is the capacitance C of a capacitor equivalent to two capacitors of respective capacitances 3.0 μF and 5.0 μF if:
    (a) They are connected in parallel?
    (b) They are connected in series?
12. If C is the capacitance equivalent to two capacitors C_1 and C_2, show that:
    (a) $C < C_1$ and $C < C_2$ if they are connected in series.
    (b) $C > C_1$ and $C > C_2$ if they are connected in parallel.
13. In Figure 22-21 find the capacitance equivalent to C_1, C_2, and C_3 connected as shown. Assume that $C_1 = C_2 = 2.0$ μF and $C_3 = 6.0$ μF.

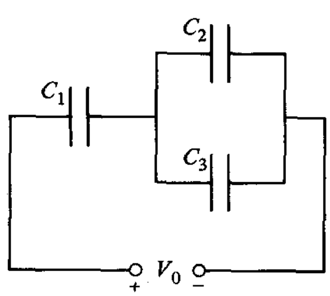


Figure 22-21

14. Suppose a potential difference of 10 volts is applied across the capacitors of Problem 13.
    (a) What is the charge on C_1?
    (b) What is the potential difference across C_1?
    (c) What is the charge on C_2 and C_3?
15. Consider the five capacitors connected as shown in Figure 22-22.

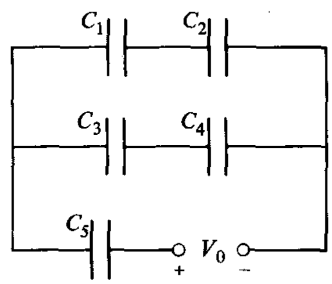


Figure 22-22

Assuming that $C_1 = 2$ μF, $C_2 = 4$ μF, $C_3 = 6$ μF, $C_4 = 8$ μF, and $C_5 = 10$ μF, calculate the capacitance C of the equivalent capacitor.
16. Suppose that for the network described in Problem 15, $V_0 = 100$ volts. **(a)** How much charge passes through the positive terminal of the source? **(b)** What is the charge on each of the five capacitors? **(c)** What is the potential across each capacitor?
17. Consider the five capacitors as arranged in Figure 22-23. Assume that $C_1 = C_3 = 2$ μF, $C_2 = C_4 = 6$ μF, $C_5 = 5$ μF, and $V_0 = 10$ volts. **(a)** Calculate the charge on each capacitor. **(b)** Find the capacitance of a single capacitor equivalent to the given arrangement.

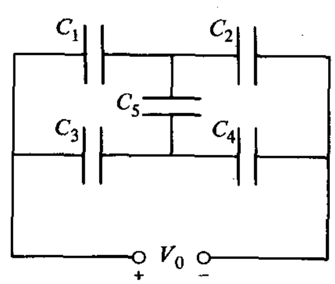


Figure 22-23

*18. Repeat Problem 17, but assume this time that $C_i = i$ μF ($i = 1, 2, 3, 4, 5$).
19. You are given a set of four 3-μF capacitors.
    (a) What equivalent capacitances can be obtained by connecting together two of these?
    (b) What equivalent capacitances

can be obtained by connecting together three of them?
(c) What equivalent capacitances can be obtained by using all four?

20. A 2-μF and a 6-μF capacitor are connected in parallel across a 5-volt line. (**a**) What energy is stored in the electrostatic field of the capacitors? (**b**) If the capacitors are disconnected from the line and reconnected with the positive plate of each connected to the negative plate of the other, what is the final value for the energy stored in the capacitors?

21. A 2-μF and a 6-μF capacitor are connected in series across a potential difference of 10 volts.
(**a**) How much energy is stored in each capacitor?
(**b**) Suppose that they are now disconnected from the source and reconnected in parallel. What is the final charge on each capacitor?
(**c**) How much energy is now stored in the capacitors?
(**d**) Compare your answers to (a) and (c) and explain, in physical terms, any differences.

22. Consider a parallel-plate capacitor of area A and separation distance d and which contains two dielectrics of constants κ_1 and κ_2, as shown in Figure 22-24. Calculate the capacitance of this capacitor. Assume that the dielectrics are identical in size.

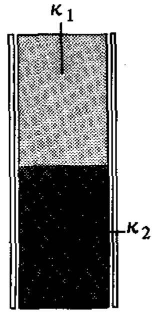


Figure 22-24

23. A parallel-plate capacitor of area A and separation distance d contains two dielectric slabs, of thicknesses a and b ($\equiv d - a$) of area A and of dielectric constants κ_1 and κ_2, respectively; see Figure 22-25. Calculate its capacitance. (*Hint*: Use (22-16) and the method of Example 22-10.)

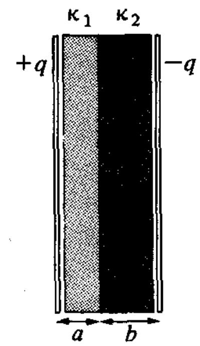


Figure 22-25

24. Consider the capacitor in Figure 22-25. Making use of (22-18) calculate (**a**) the charge density on the surfaces of the dielectrics nearest the plates and (**b**) the charge density at the interface between the dielectrics, and show that the total induced charge vanishes.

25. Show that if a dielectric slab of constant κ is introduced into the region between the plates of a parallel-plate capacitor whose plates have charge per unit area σ, then

$$\frac{\sigma_p}{\sigma} = 1 - \frac{1}{\kappa}$$

where σ_p is the induced charge on a face of the slab.

26. Show that the energy per unit volume u stored in a parallel-plate capacitor of area A and separation distance d and containing a dielectric slab of constant κ is

$$u = \frac{\kappa\epsilon_0}{2} E^2$$

where **E** is the electric field between the plates.

27. Show that the formula in Problem 26 is also applicable to a spherical capacitor. (*Note*: $|\mathbf{E}|$ is not constant in this case but varies with radius. Indeed, it can be shown that the result of Problem 26 is very generally valid for any electric field.)
28. Suppose that the plates of the capacitor in Figure 22-15 have area A, and the slab of thickness t is a conductor.
    (a) Show that the capacitance is

$$C = \frac{\epsilon_0 A}{(d - t)}$$

    (b) What charge is induced on each face of the slab if the capacitor has a charge Q?
    (c) Is a force required to move the slab closer to one of the plates? Explain.
29. Making use of (22-17), repeat (b) and (c) of Problem 28. Assume this time, however, that the slab is a dielectric of constant κ.
30. A spherical capacitor of inner radius a and outer radius b contains a concentric spherical shell of matter of inner radius c and outer radius d ($a < c < d < b$). (a) Calculate its capacitance if the shell is made of copper. (b) What would the capacitance be if the copper shell were connected to the outer sphere?
31. Repeat (a) of Problem 30 if the spherical shell is a dielectric of constant κ instead of copper. Does your answer reduce to the expected value for the special case $(d - c) = (b - a)$?
32. Assume that the entire region between the spheres in Figure 22-6 contains a dielectric of constant κ. (a) What is its capacitance? (b) If the total charge on the inner conductor is Q, what is the electric field everywhere?
33. Consider two very long, coaxial, conducting cylinders of radii a and b ($>a$), respectively, and of length L. Assuming that $L \gg a$ and b, show that the capacitance is given by (22-5). (*Hint*: Determine the electric field in the region between the cylinders by the use of Gauss' law, assuming that the capacitor has a charge Q.)

*34. A particle of charge q is at the center of a spherical dielectric shell of constant κ and radii a and b ($>a$). By applying Gauss' law and (22-16), show that the electric field E at a distance r from the particle is

$$E = \frac{q}{4\pi\epsilon_0}\frac{1}{r^2}\left\{\begin{array}{ll} 1 & r < a \\ \frac{1}{\kappa} & a \leqslant r \leqslant b \\ 1 & r > b \end{array}\right\}$$

*35. For the dielectric sphere in Problem 34, calculate (a) the charge density induced on the inner surface of the sphere and (b) the induced charge density on the outer surface, and show that the total induced charge vanishes.

36. Show that the external force required to keep the plates of a parallel-plate capacitor at the separation distance d is U/d, where U is the energy stored in the capacitor.

*37. A parallel-plate capacitor of area A and separation distance y is connected to a fixed potential difference V_0. Suppose the separation distance is increased by an external force f from y to $(y + dy)$.
    (a) Show that the change in the electrostatic energy dU is

$$dU = -\frac{1}{2}\frac{\epsilon_0 A}{y^2} V_0^2\, dy$$

    (b) Show that the charge dQ transported through the positive ter-

minal of the source is

$$dQ = -\frac{\epsilon_0 A}{y^2} V_0 dy$$

(c) What work dW_s is carried out by the source in this process?
(d) Write down the conservation of energy relation and thus show that $f = U/y$.

***38.** A parallel-plate capacitor of separation distance d and area A has a length b and a width a, so that $A = ab$. Suppose that a slab of dielectric constant κ is placed a distance y into the region between the plates; see Figure 22-18.
(a) Show that the capacitance C is

$$C = \frac{\epsilon_0}{d}[A + ay(\kappa - 1)]$$

(b) Calculate the energy $U(y)$, assuming a fixed charge Q on the capacitor.
(c) Show that the force f required to push the dielectric a distance dy is

$$f = -\frac{Q^2}{2\epsilon_0}\frac{(\kappa - 1)\,ad}{[A + ay(\kappa - 1)]^2}$$

***39.** Repeat all parts of Problem 38, but assume this time that the potential V_0, across the capacitor is kept fixed.

40. In Problem 38 calculate the work W

$$W = \int_0^b F(y)\,dy$$

required to insert the dielectric completely. Compare this with the energy change in the electrostatic field during this process.

***41.** Consider four parallel metal plates, each of area A, and spaced a distance d apart so that the distance between the first and last plate is $3d$. Show that if alternate plates are connected by conducting wires the capacitance C is

$$C = \frac{3\epsilon_0 A}{d}$$

(*Hint*: Why, if q is the charge on the first plate, must the charge on the next be $-2q$?)

†42. For the physical situation described in Problem 34, show that the displacement vector **D** is everywhere radial and has the magnitude

$$D = \frac{q}{4\pi r^2}$$

†43. Consider, in Figure 22-26, a rectangular slab of a dielectric in which there is a *uniform* dipole moment per unit volume **P** directed perpendicular to a face.

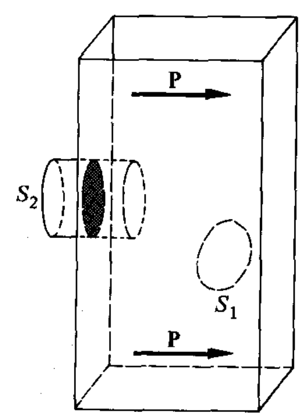


Figure 22-26

(a) Show that if S_1 is a closed surface contained entirely within the dielectric, then the total induced charge inside, $q_p(S_1)$, vanishes. (*Hint*: Use (22-25) and the fact that $\oint d\mathbf{S} = 0$ for any closed surface.)
(b) By making use of the closed surface S_2 in the figure, show that the polarization charge σ_p on the dielectric surface is

$$\sigma_p = -P.$$

23 The electric current

Nothing is too wonderful to be true if it be consistent with the laws of nature and in such things as this experiment is the best test of such consistency.

MICHAEL FARADAY (1791–1867)

23-1 Introduction

The preceding four chapters have dealt mainly with physical situations involving *static* distributions of charge. We now turn from these considerations to study a completely different set of phenomena, which are associated with electric charge in motion. In this connection an *electric current* or a *current* will be said to exist in a certain region of space if macroscopic amounts of charge undergo ordered motion in that region. As an introduction to a study of phenomena associated with electric currents, in this and the following chapter we first consider the problem of their generation, measurement, and control.

Figure 23-1 shows how a current may be generated very simply by use of a charged capacitor. Part (a) depicts the initial situation of a capacitor of charge Q and some of the electric-field lines associated with it. Suppose that these conductors are suddenly connected by a metallic wire. Because of the existence of a potential difference between the conductors, at least initially, the electric field in the interior of the connecting wire is *not* zero

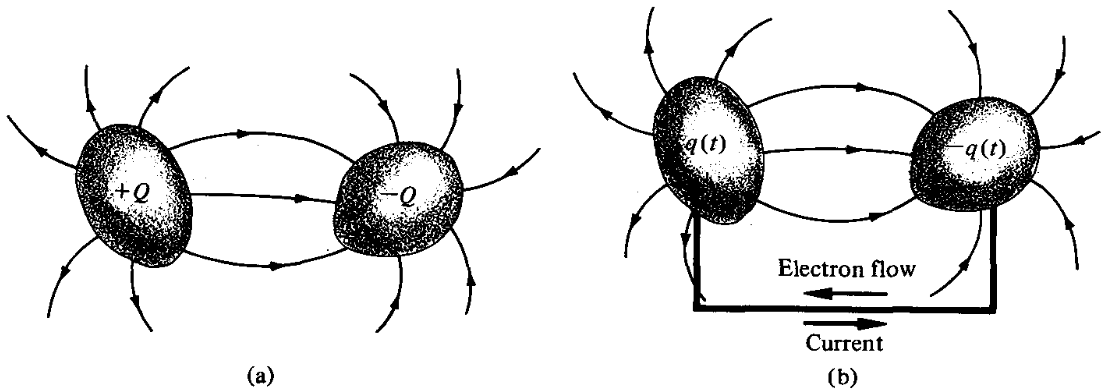


Figure 23-1

and thus electrons will start to flow from the negative conductor to its positively charged partner. As a result of this flow, the charge on the capacitor will start to decrease and will continue to do so as long as any charge remains on the conductors. According to the above definition, this flow of electrons along the connecting wire in Figure 23-1b constitutes an electric current. Because of its generally short-lived nature, this particular current is called a ***transient*** current.

Mainly for historical reasons, it is customary to define electric current, not in terms of the actual flow of negatively charged electrons, but rather in terms of an equivalent flow of positive charge in the opposite direction. Thus, even though the electrons travel along the connecting wire from right to left in Figure 23-1b, the electric current is by definition said to be directed in the opposite sense from left to right. This choice is arbitrary but the one conventionally made.

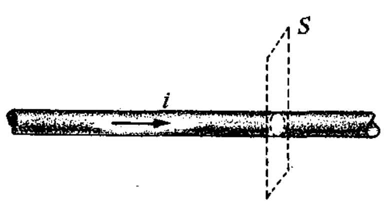


Figure 23-2

Consider, in Figure 23-2, a short segment of a wire through which flows a certain electric current. Let us assign a sense of direction to this current by arbitrarily drawing an arrowhead directed, say, to the right. Also, define S to be a surface whose plane is perpendicular to the current and thus to the wire at some point. Let Δq represent the charge that during an infinitesimal time interval Δt crosses S in the assumed direction of the current flow, that is, to the right in the figure. The current i passing through S is then defined by the limit

$$i = \lim_{\Delta t \to 0} \frac{\Delta q}{\Delta t}$$

which, according to the definition of a derivative, is the same as

$$i = \frac{dq}{dt} \tag{23-1}$$

Note that i can be positive or negative depending on the sign of Δq. A negative value for i means that the current flows in a direction opposite to that assumed in the figure. Thus, $i < 0$ means that during the time interval Δt either a certain amount of positive charge moves through S directed to the left in the figure or, equivalently, negative charge moves to the right. Corresponding statements apply for positive values for i.

It follows from the definition in (23-1) that the coulomb per second is the unit of current. The SI unit of current is called the *ampere* (abbreviated A), and is given by

$$1\ \text{A} \equiv 1\ \text{ampere} = 1\ \text{C/s} \tag{23-2}$$

Two units related to this are the milliampere (mA) and the microampere (μA), which are defined to be 10^{-3} ampere and 10^{-6} ampere, respectively.

Example 23-1 Suppose that the charge $q(t)$ on the capacitor in Figure 23-1b varies in time as

$$q(t) = Qe^{-t/\tau}$$

with $Q = 3.0\ \mu\text{C}$ and $\tau = 1.0\ \mu\text{s}$. Calculate the current i in the connecting wire at any time t.

Solution Let us assume that, contrary to the actual situation, the direction of the current in the connecting wire is to the *left* in the figure. Then the amount of charge Δq that crosses a surface perpendicular to the connecting wire in a time interval Δt is the same as the *increase* of the charge on the capacitor. Hence, if $q(t)$ is the charge on the capacitor at time t, then, according to (23-1),

$$\begin{aligned} i &= \frac{dq}{dt} = \frac{d}{dt}(Qe^{-t/\tau}) = -\frac{Q}{\tau}e^{-t/\tau} \\ &= -3.0e^{-t/\tau} \quad \text{(A)} \end{aligned}$$

where we have used the relation $d(e^{\alpha t})/dt = \alpha e^{\alpha t}$ and in the last equality the given values for Q and τ. The minus sign here reflects the fact that contrary to the stated assumption, but consistent with our physical expectations, the charge flow along the wire is directed to the right in the figure; that is, from the positive to the negative plate of the capacitor.

23-2 Electromotive force

The current associated with the capacitor discharge in Figure 23-1 is a *transient* since it exists only for the short period of time that the capacitor has a charge. To produce a *steady current*—that is, one which persists for

long periods of time—it is necessary to use a different approach. The purpose of this section is to describe, in general terms, a class of devices that can be used to produce steady currents.

In the following, by the term *battery* we shall mean any one of a number of devices that are capable of maintaining a fixed potential difference between two points, and thus of producing a steady current between them. The terms a *source* or a *seat of electromotive force* are also in common usage. A complete listing of batteries would include chemical batteries, solar cells, thermoelectric cells, and electromagnetic generators, where in each case the preceding adjective describes the source of the energy involved. Thus the energy output of a chemical battery results from the chemical processes that take place in its interior, and the output of a solar cell is the result of radiant energy being incident on—and thereby interacting with—the material of the cell. Of these, we shall examine in detail only the electromagnetic generator. This device, as will be discussed in Chapter 27, makes possible the direct conversion of mechanical energy of motion into electrical energy.

An essential and visible feature of any battery, regardless of its type, is a pair of conducting terminals, which are outside the battery and are often labeled plus (+) and minus (−). As a general rule, the plus terminal has a positive charge and is therefore at a higher potential than is the other terminal with its negative charge. These terminals are always connected *inside the battery* by a conducting path of some type, such as a metallic strip or an electrolytic bath. If no conductor connects the terminals *outside* the battery, then the battery is said to be on *open circuit*. In general, no current flows inside or outside of a battery when it is on open circuit.

Figure 23-3 shows the external essentials of a battery on open circuit, with its plus and minus terminals labeled A and B, respectively. As is implied by the electric-field lines in the figure, there is a potential difference V_{AB} between these terminals. According to (21-2) and (21-5), V_{AB} may be expressed in the form

$$V_{AB} \equiv V_A - V_B = -\int_B^A \mathbf{E} \cdot d\mathbf{l} = \int_A^B \mathbf{E} \cdot d\mathbf{l} \qquad \textbf{(23-3)}$$

where, for example, V_A is the potential of the plus terminal and $\mathbf{E}$ is the electric field evaluated at points along the path connecting the terminals. As we saw in Chapter 21, the electrostatic field is conservative, and thus this

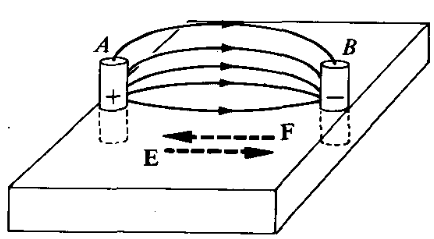


Figure 23-3

integral is independent of the path followed in going from A to B. This path may be entirely outside the battery, entirely inside, or partially inside and partially outside.

Consider now the situation *inside* a battery on open circuit. Because of the potential difference between the terminals, and the existence inside the battery of a conducting path between them, it follows that inside the battery there must exist a force field $\mathbf{F}$ that is distinct from the electric field $\mathbf{E}$ produced by the charge at the terminals. Otherwise, in response to the electric field, current would flow inside the battery and neutralize the charge on the two terminals. In particular, then, since on open circuit equilibrium conditions prevail, this force $\mathbf{F}$ on an electron or ion of charge q_0 must be related to the electric field $\mathbf{E}$ *inside* the battery by

$$\mathbf{F} + q_0\mathbf{E} = 0 \tag{23-4}$$

It is convenient to characterize this force field $\mathbf{F}$ inside a battery in terms of a certain quantity called the *electromotive force* or *emf* $\mathscr{E}$ of the battery. The emf $\mathscr{E}$ is defined to be the work carried out by the force field $\mathbf{F}$ in taking a particle of unit positive charge from the minus to the plus terminal *inside* the battery. Thus

$$\mathscr{E} = \frac{1}{q_0}\int_B^A \mathbf{F} \cdot d\mathbf{l}$$

and therefore the product $q\mathscr{E}$ represents the work carried out by a battery in transporting a charge q between its terminals. On substituting into this relation, by use of (23-3) and (23-4), we find that

$$\mathscr{E} = V_{AB} \qquad \text{(open circuit)} \tag{23-5}$$

so the emf $\mathscr{E}$ of a battery is precisely the potential difference between its terminals on *open circuit*.

It is important, despite the equality in (23-5), to keep the physical difference between $\mathscr{E}$ and V_{AB} in mind. The former represents the work per unit charge carried out by the internal energy source of the battery, whereas the latter is the potential difference between the terminals of the battery due to charge separation. The validity of (23-5) depends on (23-4), and therefore holds only under conditions of open circuit.

According to the above definition, the unit of electromotive force is the same as that of potential. Thus the volt is the unit of emf and when speaking of a "10-volt battery" what is meant is one that, on open circuit, will sustain a potential difference of 10 volts between its terminals.

Let us now consider briefly what happens to a battery not on open circuit. Imagine connecting a conducting wire between the terminals in Figure 23-3. Because of the initial potential difference V_{AB} between these terminals, current will flow along this wire from A to B. Associated with this current there will be a momentary decrease of charge on the battery terminals. Thus the electric field $\mathbf{E}$ decreases inside the battery and (23-4) ceases to be valid.

Instead, the strength of the force field **F** now exceeds that of the electric force, so that effectively positive charge is forced to go in the direction from B to A *inside* the battery. In other words, a current flows inside the battery from the negative to the positive terminals. Thus there is a current from A to B along the external conducting path and one from B to A inside the battery. Under steady-state conditions these two currents will be identical, so a continuous stream of charge will be going around a closed path. Unfortunately, the relation in (23-5) between the emf of the battery and the potential difference between its terminals is not generally valid under these conditions. However, assuming that there are no other batteries in the circuit, we find that the difference ($\mathscr{E} - V_{AB}$) is generally very small. Hence it will usually be neglected. A battery for which the difference ($\mathscr{E} - V_{AB}$) is significant is said to have an *internal resistance.*

23-3 Drift velocity and current density

Under conditions of static equilibrium there can be no electric field in the interior of a conductor. The free electrons inside behave as a dilute gas and wander through the positive lattice in such a way that from a macroscopic viewpoint the conductor is electrically neutral. The density of these electrons is very high, and in ordinary metals at room temperature it is of the order of 10^{29} per cubic meter. Just as for an ordinary gas, we may think of these electrons as being in a state of constant motion, with a velocity that typically is of the order of 10^6 m/s. The velocity distribution of these electrons is isotropic, so at any point in the lattice the average flow of electrons in any given direction is compensated for by an equal flow in the opposite direction. This means that under normal circumstances when the electric field inside the metal vanishes there is no electric current there either.

Let us now contrast this picture with the corresponding one when the electric field inside the conductor is not zero. Figure 23-4 depicts the motion of a typical electron in a metal when there is an electric field present. Assuming this field acts to the left, this electron will accelerate to the right. However, before it can pick up very much speed, it will collide with an ion,[1] and as a result it will, in general, lose energy and change its direction of motion. After this collision, it will again accelerate to the right and continue to do so until it suffers a second inelastic collision with an ion at some point b. As shown in the figure, the motion of the electron thus consists of a zigzag pattern of short, segmented trajectories *abcdefgh*. However, this pattern is not random. Because of the electric field, the electrons tend to drift

[1] According to the ideas of quantum mechanics, an electron traveling through a "perfect" lattice will not collide with a lattice ion. However, such electrons may undergo collisions with lattice imperfections, such as impurities or vacancies, and all text references to collision with ions should be interpreted in this sense.

gradually to the right. In other words, because of the ions, the electrons do not accelerate to arbitrarily high velocities, but rather when viewed over long periods of time they seem to move gradually in a direction opposite to that of the electric field. The velocity associated with this motion of the electrons is called the *drift velocity* and will be represented by the symbol $\mathbf{v}_d$. We shall see below that even though the velocity distribution of the electrons is isotropic and peaks at a value of the order of 10^6 m/s, the drift velocity $\mathbf{v}_d$ is directed opposite to the direction of the field and is, in general, very much smaller.

Let us estimate the drift velocity in a typical case. To this end, consider, in Figure 23-5, a segment of a wire of cross-sectional area A in which there flows a uniform current i. In a small time interval Δt, the charge ΔQ that flows past a given surface S is that contained in a cylinder of cross-sectional area A and length $v_d\,\Delta t$. That is, since we view the electrons as moving at the drift velocity $\mathbf{v}_d$, all electrons in this region of volume $Av_d\,\Delta t$ will cross the surface S and thus contribute to the current. If n represents the number of electrons per unit volume, it follows that the net charge ΔQ, which flows through S along the direction of the current, is

$$\Delta Q = en(Av_d\,\Delta t)$$

where e is the magnitude of the elementary quantum of charge and has the value 1.6×10^{-19} coulomb. Dividing both sides by Δt and equating the current i to the ratio $\Delta Q/\Delta t$, we thus obtain

$$i = neAv_d \tag{23-6}$$

Equivalently, this may be reexpressed in the form

$$\mathbf{j} = -ne\mathbf{v}_d \tag{23-7}$$

where $\mathbf{j}$ is a vector *along* the direction of the current and is known as the *current density*. It has the magnitude

$$j = \frac{i}{A} \tag{23-8}$$

The minus sign in (23-7) is necessary since, by our convention, the direction of the current is opposite to the drift velocity $\mathbf{v}_d$ of the electrons.

As a typical case, consider a wire of cross-sectional area 1 mm^2 = 10^{-6} m^2

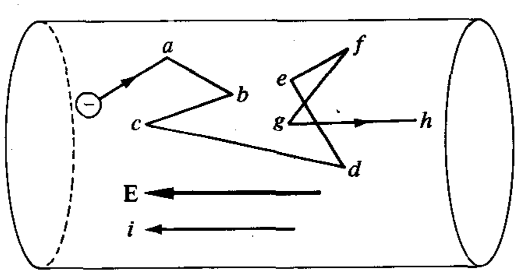


Figure 23-4

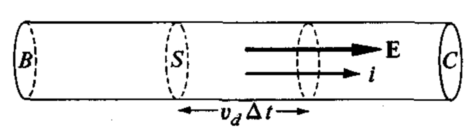


Figure 23-5

through which flows a current of 1.0 ampere. Assuming that $n \cong 10^{29}/\text{m}^3$, on solving (23-6) for v_d we obtain

$$v_d = \frac{i}{neA} = \frac{1.0\ \text{A}}{(10^{29}/\text{m}^3)\times(1.6\times10^{-19}\ \text{C})\times10^{-6}\ \text{m}^2}$$
$$= 6.3\times10^{-5}\ \text{m/s}$$

As anticipated above, this is a very small velocity compared to the value 10^6 m/s, which characterizes the velocity of the electrons in a metal.

Example 23-2 An aluminum wire has a density of 2.7 g/cm^3 and a cross-sectional area of 1.0×10^{-5} m^2, and carries a current of 2.0 ampere.

(a) What is the current density in the wire?

(b) Assuming three free electrons per atom, calculate the density of electrons in the wire.

(c) What is the drift velocity of these electrons?

Solution

(a) Making use of the definition for current density in (23-8) and the given values for i and A, we find

$$j = \frac{i}{A} = \frac{2.0\ \text{A}}{10^{-5}\ \text{m}^2} = 2.0\times10^5\ \text{A/m}^2$$

(b) Since the atomic mass of aluminum is 27, 1 mole, or 27 grams of this material, contains 6.0×10^{23} atoms. Further, since the density of the aluminum wire is 2.7 g/cm^3, it follows that 10 cm^3 of aluminum contains 1 mole. In other words, 1 cm^3 of aluminum contains 6.0×10^{22} atoms so there are altogether 6.0×10^{28} atoms/m^3. With three free electrons per atom, it follows then that the electron density is

$$n = 1.8\times10^{29}/\text{m}^3$$

(c) On solving (23-7) for v_d and substituting the values for the various parameters we obtain

$$v_d = \frac{j}{ne} = \frac{2.0\times10^5\ \text{A/m}^2}{(1.8\times10^{29}/\text{m}^3)\times(1.6\times10^{-19}\ \text{C})}$$
$$= 6.9\times10^{-6}\ \text{m/s}$$

23-4 Ohm's law

We have seen in Section 23-3 that in the presence of an electric field the electrons in a metal gradually drift in a direction opposite to that of the field with a certain drift velocity $\mathbf{v}_d$. Physically, we might expect that the larger the electric field, the greater will be this drift velocity. Experiment shows that this is indeed the case and that the precise relationship between $\mathbf{v}_d$ and $\mathbf{E}$ is given by a certain empirical relation known as *Ohm's law*. The purpose of this section is to discuss this law.

Consider a region inside a conductor in which there is an electric field $\mathbf{E}$. Following convention, we shall continue to interpret the associated electron

flow opposite to the field in terms of an equivalent flow of positive charge along **E**. As for the case of the uniform current in Section 23-3, it is convenient to characterize this flow by a vectorial current density **j** defined so that the quantity $\mathbf{j} \cdot d\mathbf{S}$ represents the flow of charge per unit time across any area element $d\mathbf{S}$ inside the conductor. Since the quantity **j** represents a flow of charge and since the cause of this flow is the electric field **E** itself, we might expect a relation of some type to exist between them. Indeed, experiment shows that, for many conductors, **j** and **E** are simply proportional to each other. For these substances we may write

$$\mathbf{j} = \sigma \mathbf{E} \tag{23-9}$$

a relation known as *Ohm's law*. The coefficient of proportionality, σ, is a constant independent of **E** and is called the *conductivity*.

It should be noted that since it is not feasible to measure **j** and **E** separately inside a conductor, Ohm's law in the form of (23-9) cannot be verified directly. We shall derive below an alternate form of this law, which involves quantities more amenable to experimental measurement. It is this latter form that is used almost exclusively in most applications.

It is convenient to define a unit of electrical resistance called the *ohm* (abbreviated Ω) by the relation

$$1 \text{ ohm} \equiv 1\ \Omega = 1\ \text{V/A}$$

Using the fact that the unit of electric field is the volt per meter (V/m) and that of current density is the ampere per square meter (A/m^2), we find by use of (23-9) that the unit of conductivity is $(\Omega\text{-m})^{-1}$. A quantity related to the conductivity of a material is its *resistivity* ρ. This is defined to be the reciprocal of the conductivity σ:

$$\rho = \frac{1}{\sigma} \tag{23-10}$$

It follows that the Ω-m is the appropriate unit of resistivity.

Table 23-1 lists the conductivities and the associated resistivities of typical metals. Note that Ag and Cu have relatively high conductivities and thus are better conductors than are the other elements listed. Physically, this means that for a given electric field strength, more current will flow in a silver or copper wire than in an iron wire of the same cross-sectional area. As will be

Table 23-1 Conductivities and resistivities at 293 K

Substance	$\sigma\ (\Omega\text{-m})^{-1}$	$\rho\ (\Omega\text{-m})$
Ag	6.3×10^{7}	1.6×10^{-8}
Al	3.6×10^{7}	2.8×10^{-8}
Cu	5.9×10^{7}	1.7×10^{-8}
Fe	1.0×10^{7}	1.0×10^{-7}
Ni	1.4×10^{7}	7.1×10^{-8}
W	1.8×10^{7}	5.6×10^{-8}

described in Section 23-8, σ also depends on temperature, and for most substances the higher the temperature the lower is the conductivity.

To cast Ohm's law into a more useful form, consider the wire in Figure 23-5 and suppose that it has a conductivity σ, a length l, and a cross-sectional area A. Assuming that the current i is uniform across the wire, we shall establish below that the potential difference $V \equiv V_B - V_C$ between its end faces is

$$V = Ri \tag{23-11}$$

where R is a parameter known as the *resistance* of the wire and is defined by

$$R = \frac{l}{A\sigma} \tag{23-12}$$

In words, the relation in (23-11), which is also known as Ohm's law, states that the ratio of the potential drop V between the ends of a wire to the current i flowing in it, is constant, independent of V and i. According to (23-12), this ratio R varies directly as the length l of the wire and inversely with the cross-sectional area A.

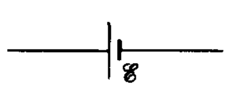
To establish (23-11), note first that the potential difference $V \equiv (V_B - V_C)$ between the end faces in Figure 23-5 may be expressed as a line integral of the electric field $\mathbf{E}$ along a straight line parallel to the current flow. Thus

$$V = V_B - V_C = \int_B^C \mathbf{E} \cdot d\mathbf{l} = \int_B^C \frac{1}{\sigma} \mathbf{j} \cdot d\mathbf{l}$$
$$= \frac{j}{\sigma} \int_B^C dl = \frac{jl}{\sigma}$$

where the third equality follows from (23-9) and the fourth from the fact that the current flow has been assumed to be uniform. The validity of (23-11), with R defined in (23-12), now follows by use of (23-8).

Just as for a capacitor, it is convenient to have available a symbol to denote a resistor in a circuit. Following convention we shall use the symbols

the first of which represents a resistor of resistance R and the second a variable resistor. Similarly, we shall use either of the symbols



to represent a battery of emf $\mathscr{E}$. The positive terminal is in each case associated with the left vertical line and the negative terminal with the shorter vertical line to the right.

Example 23-3 An aluminum wire has a cross-sectional area of 2.0 mm^2 and a length of 30 meters. Using the data in Table 23-1, calculate its resistance.

Solution Substituting into (23-12) and using the value $\sigma = 3.6 \times 10^7\ (\Omega\text{-m})^{-1}$, we find that

$$R = \frac{l}{A\sigma} = \frac{30\ \text{m}}{(2.0 \times 10^{-6}\ \text{m}^2) \times [3.6 \times 10^7\ (\Omega\text{-m})^{-1}]}$$
$$= 0.42\ \Omega$$

Example 23-4 A battery of emf $\mathscr{E} = 10$ volts is connected across a 15-Ω resistor, as in Figure 23-6. What current flows?

Solution According to (23-5), the emf of the battery is equal to the potential drop across the resistor. But the latter is *Ri*, according to (23-11). Thus

$$\mathscr{E} = Ri$$

or

$$i = \frac{\mathscr{E}}{R} = \frac{10\ \text{V}}{15\ \Omega} = 0.67\ \text{A}$$

The direction of this current through the resistor is, as shown in the figure, from the plus to the minus terminal of the battery.

Example 23-5 Suppose that a current of 2.0 amperes flows clockwise in the circuit in Figure 23-7. If V_f, the potential at point *f*, is zero, calculate the following potentials: (a) V_a; (b) V_b; (c) V_c; (d) V_d; and (e) V_e.

Solution

(a) Since the emf of the lower battery is 50 volts and since $V_f = 0$, it follows that

$$V_a - V_f = V_a = 50\ \text{V}$$

(b) Since there is no resistance from *a* to *b* (straight lines on circuit diagrams always represent resistance-free paths) it follows that

$$V_b = V_a = 50\ \text{V}$$

(c) A 2.0-ampere current flowing through a 5-Ω resistor produces, according to (23-11), a potential drop of 10 volts. Thus, $V_b - V_c = 10$ volts or, in other words,

$$V_c = V_b - 10\ \text{V} = 50\ \text{V} - 10\ \text{V} = 40\ \text{V}$$

(d) Because of the 20-volt battery,

$$V_c - V_d = 20\ \text{V}$$

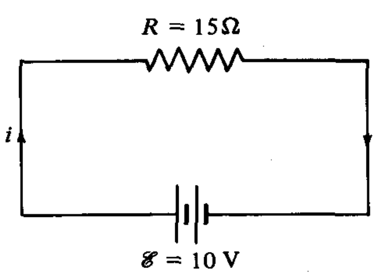


Figure 23-6

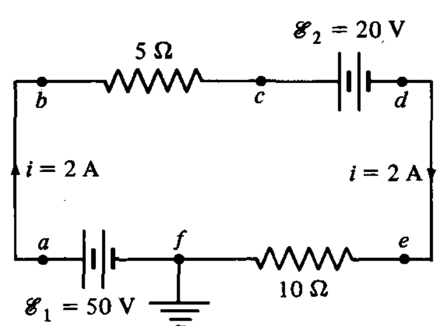


Figure 23-7

Thus

$$V_d = V_c - 20\text{ V} = 40\text{ V} - 20\text{ V} = 20\text{ V}$$

(e) As in (b), $V_d = V_e$ and thus $V_e = 20$ volts.

It is interesting to note that the current through the upper battery ($\mathscr{E}_2$ in the figure) is directed the "wrong" way; that is, the current flows from the positive to the negative terminal *inside* the battery. Under these circumstances we say that the battery is *being charged.* Indeed, in this case, part of the energy output of $\mathscr{E}_1$ goes to increase that of $\mathscr{E}_2$.

23-5 Discharge of a capacitor

The purpose of this section is to reexamine the physical situation described in Section 23-1 involving the discharge of a capacitor through a resistor. This time, however, we shall analyze the charge flow quantitatively.

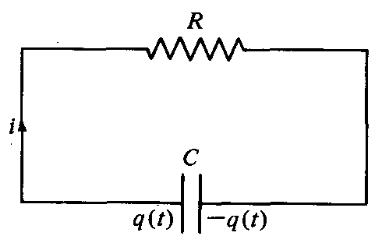


Figure 23-8

Consider, in Figure 23-8, a capacitor of capacitance C, which has an original charge Q_0 and across which a resistor of resistance R is connected. If $q \equiv q(t)$ is the charge on the capacitor at time t after the discharge has begun, then by definition of capacitance the potential across the capacitor at this instant is q/C. Assuming, as shown in Figure 23-8, that the left-hand plate has the positive charge and is thus at the higher potential, it follows that the current i in the circuit will be directed clockwise, as shown. According to (23-11), the potential difference between the ends of the resistor is Ri. Since this must be the same as the potential drop across the capacitor, it follows that

$$\frac{q}{C} = Ri \tag{23-13}$$

Finally, since a positive value for the current i means in the present case that the charge $q(t)$ on the capacitor is decreasing, that is, for $i > 0$, $dq < 0$, it follows that

$$i = -\frac{dq}{dt} \tag{23-14}$$

Substitution into (23-13) thus leads to

$$R\frac{dq}{dt} + \frac{1}{C}q = 0 \tag{23-15}$$

This relation is known as the *circuit equation* for the electric circuit in Figure 23-8. Its solution yields the charge q on the capacitor, as well as the current i in the circuit at any time t.

To solve (23-15), it is convenient first to cast it into the form

$$\frac{dq}{q} = -\frac{1}{RC}\,dt$$

which may be integrated to

$$\ln q = -\frac{t}{RC} + \alpha$$

with α an integration constant. Since at $t = 0$ the charge on the capacitor is Q_0, that is, $q(0) = Q_0$, it follows that $\alpha = \ln Q_0$, and thus

$$\ln q = \ln Q_0 - \frac{t}{RC}$$

Finally, taking the exponential of both sides and recalling that $e^{\ln q} = q$, we obtain, as the solution for (23-15),

$$q(t) = Q_0 e^{-t/RC} \tag{23-16}$$

Thus the charge on the capacitor drops exponentially to zero from its initial value Q_0. The current i in the circuit is found by substitution into (23-14) to be

$$i = \frac{Q_0}{RC}\, e^{-t/RC} \tag{23-17}$$

and thus i also drops off exponentially from its initial value of Q_0/RC. Figure 23-9 shows a plot of q and i as functions of time.

In connection with a study of the charging and the discharging of capacitors, it is convenient to define a *time constant* τ, which for the case just considered represents the time it takes the charge on the capacitor to drop to $1/e$ ($\cong 0.368$) of its initial value. Thus τ satisfies

$$q(\tau) = \frac{1}{e}\,Q_0$$

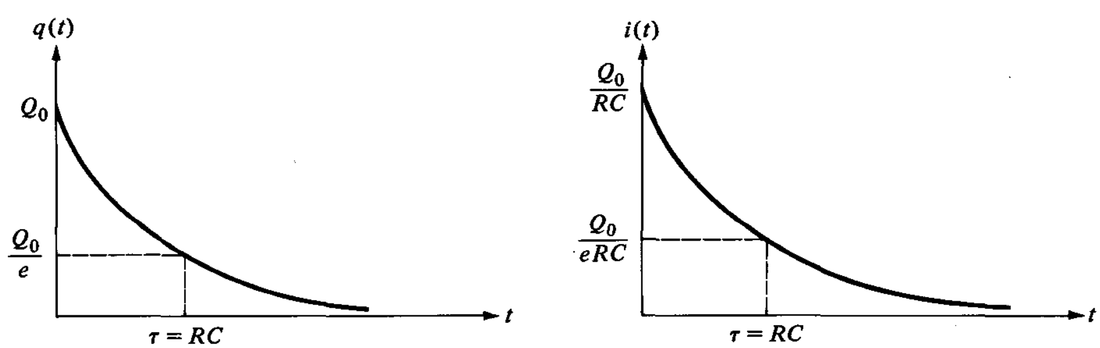


Figure 23-9

and, substituting for $q(\tau)$ from (23-16), we obtain

$$\tau = RC \tag{23-18}$$

Example 23-6 What is the time constant τ associated with the discharge of a 3.0-μF capacitor through a 100-Ω resistor?

Solution Substituting the given values for R and C into (23-18), we find that

$$\tau = RC = 100\,\Omega \times 3.0 \times 10^{-6}\,\text{F} = 3.0 \times 10^{-4}\,\text{s}$$

where we have used the fact that the unit of the (Ω-F) is the same as the second. Thus in a time interval of 3.0×10^{-4} second the charge on the capacitor drops from its original value, whatever it may have been, to $1/e \cong 37$ percent of this value.

Example 23-7 A 2.0-μF capacitor is charged by a battery of emf $\mathscr{E} = 10$ volts. It is then disconnected from the battery and shorted out by a $1.0 \times 10^3\ \Omega$ resistor.
(a) What is the maximum charge on the capacitor?
(b) What is the time constant?
(c) What is the maximum current flowing in the resistor?
(d) How much charge is left on the capacitor 1.0 second after the discharge starts?

Solution
(a) According to (23-5), the maximum charge Q_0 on the capacitor is

$$Q_0 = C\mathscr{E} = (2.0 \times 10^{-6}\,\text{F}) \times (10\,\text{V}) = 2.0 \times 10^{-5}\,\text{C}$$

(b) The substitution of the given values for R and C into (23-18) yields

$$\tau = RC = (1.0 \times 10^3\,\Omega) \times (2.0 \times 10^{-6}\,\text{F}) = 2.0 \times 10^{-3}\,\text{s}$$

(c) The maximum current, according to (23-17), flows at $t = 0$, and thus

$$\begin{aligned} i_{\text{max}} = i(0) &= \frac{Q_0}{RC} = \frac{Q_0}{\tau} = \frac{2.0 \times 10^{-5}\,\text{C}}{2.0 \times 10^{-3}\,\text{s}} \\ &= 1.0 \times 10^{-2}\,\text{A} \end{aligned}$$

where in the fourth equality we have used the value for Q_0 and τ from (a) and (b), respectively.
(d) On setting $t = 1$ second in (23-16), we obtain for $q(1\text{ s})$

$$\begin{aligned} q(1\text{ s}) &= Q_0 \exp\left\{-\frac{1\text{ s}}{\tau}\right\} = (2.0 \times 10^{-5}\,\text{C}) \times \exp\left\{-\frac{1\text{ s}}{2.0 \times 10^{-3}\,\text{s}}\right\} \\ &= 2.0 \times 10^{-5} \times e^{-500}\,\text{C} \\ &\approx 10^{-200}\,\text{C} \end{aligned}$$

For practical purposes, then, the capacitor is completely discharged after 1 second. More realistically, after a capacitor has been discharging for a time interval of the order, say, of 10τ, it is essentially discharged.

23-6 The charging of a capacitor

Having seen how to describe quantitatively the discharge of a capacitor, in this section we examine the corresponding problem of the charging of a capacitor by use of a battery.

Consider, in Figure 23-10, a capacitor of capacitance C in series with a resistor R, both connected across a battery of emf $\mathscr{E}$. For convenience we have also included in the circuit a switch S and we shall assume that at $t = 0$ the switch is closed.

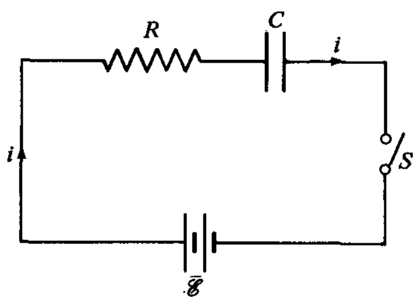


Figure 23-10

To obtain the circuit equation, assume, as shown in the figure, that the current flows in a clockwise sense. The potential difference across the terminals of the battery is the sum of the potential drops Ri and q/C across the resistor and the capacitor, respectively. According to (23-5), we have then

$$\mathscr{E} = \frac{q}{C} + Ri$$

This time, since the capacitor is charging up, a positive value for i means that $dq > 0$. Hence the current i is related to the capacitor charge q by (23-1), and substitution into the above formula leads to the circuit equation

$$\mathscr{E} = \frac{q}{C} + R\frac{dq}{dt} \tag{23-19}$$

Note that for $\mathscr{E} = 0$ this reduces to (23-15), as it should.

To solve (23-19), let us reexpress it as

$$\frac{dq}{\mathscr{E}C - q} = \frac{dt}{RC}$$

and then integrate both sides. The result is

$$\ln(\mathscr{E}C - q) = -\frac{t}{RC} + \alpha$$

and since at $t = 0$, $q = 0$, the integration constant α is found to be $\ln(\mathscr{E}C)$. Solving the resulting formula for q, we obtain

$$q = \mathscr{E}C(1 - e^{-t/RC}) \tag{23-20}$$

which by use of (23-1) yields the associated formula for the current

$$i(t) = \frac{\mathscr{E}}{R} e^{-t/RC} \tag{23-21}$$

Note that at $t = 0$, when the switch is closed, the current $i(0)$ in the circuit is the same as the current that would flow if the capacitor were *not* present.

Figure 23-11 shows a plot of q and i as functions of time. Initially, the current has the value $\mathscr{E}/R$ and from there it decreases exponentially with the same time constant $\tau = RC$ as found in Section 23-5. Thus, at $t = \tau$, the current has dropped to $1/e$ of its initial value $\mathscr{E}/R$. In a similar way, the charge on the capacitor rises from its initial value zero, to within the factor

$$1 - \frac{1}{e} \approx 1 - 0.37 = 0.63$$

of its saturated value $\mathscr{E}C$ in the same time interval, of length $\tau = RC$. Thus, in a period of the order of 10τ the charge on the capacitor assumes its final value to a high degree of precision.

To summarize, if a resistor and capacitor are connected in series across a battery, as in Figure 23-10, then initially a current $i = \mathscr{E}/R$ will flow in the circuit as if the capacitor were not present. In a time interval of the order of $\tau = RC$, however, this current decreases as the charge on the capacitor increases. This buildup of charge will continue until finally enough charge has accumulated on the capacitor that its potential is the same as the emf of the battery. At this point the potential drop across the resistor must be zero, and thus current can no longer flow.

Example 23-8 A battery of emf $\mathscr{E} = 10$ volts is connected across a 3.0-μF capacitor and a 100-Ω resistor in series. Assuming that at $t = 0$ the switch is closed, calculate:

(a) The initial current in the circuit.
(b) The final charge on the capacitor.
(c) The charge on the capacitor at $t = 3.0 \times 10^{-4}$ second.
(d) The current at $t = 9.0 \times 10^{-4}$ second.

Solution

(a) The initial current i_0 has the same value as that of the current which would flow

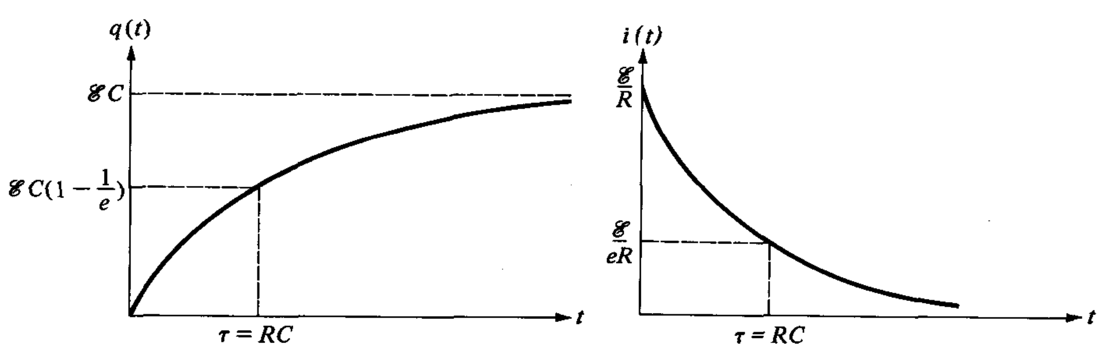


Figure 23-11

if the capacitor were not present. Thus

$$i_0 = \frac{\mathscr{E}}{R} = \frac{10\ \text{V}}{100\ \Omega} = 0.1\ \text{A}$$

(b) The maximum charge q_f on the capacitor is that value for which the potential across it is the same as the emf of the battery. Thus

$$\begin{aligned} q_f &= C\mathscr{E} = (3.0 \times 10^{-6}\ \text{F}) \times (10\ \text{V}) \\ &= 3.0 \times 10^{-5}\ \text{C} \end{aligned}$$

(c) Since $R = 100\ \Omega$ and $C = 3.0\ \mu\text{F}$, the time constant τ has the value

$$\tau = RC = 100\ \Omega \times 3.0 \times 10^{-6}\ \text{F} = 3.0 \times 10^{-4}\ \text{s}$$

Thus, according to (23-20), at time t

$$\begin{aligned} q(t) &= C\mathscr{E}(1 - e^{-t/RC}) \\ &= 3.0 \times 10^{-5}(1 - e^{-t/3\times 10^{-4}\text{s}})\ \text{C} \end{aligned}$$

and, in particular, at $t = 3.0 \times 10^{-4}$ second

$$\begin{aligned} q(3 \times 10^{-4}\ \text{s}) &= 3.0 \times 10^{-5}\,(1 - e^{-1})\ \text{C} \\ &= 1.9 \times 10^{-5}\ \text{C} \end{aligned}$$

(d) Substituting the given values for $\mathscr{E}$, R, and C we find in the same way that at $t = 9.0 \times 10^{-4}$ second $\equiv 3\tau$ the current in the circuit is

$$\frac{\mathscr{E}}{R} e^{-3\tau/\tau} = 0.1 e^{-3}\ \text{A} = 5.0 \times 10^{-3}\ \text{A}$$

23-7 Electric power—Joule heating

In Section 23-3 we noted that in the presence of an electric field the electrons in a conductor accelerate in a direction opposite to that of the field, and in the resulting collisions with the lattice ions they tend to lose energy. This energy loss by the electrons is gained by the lattice and manifests itself macroscopically by the heating up of any substance through which a current flows. This phenomenon is called *Joule heating.* The purpose of this section is to obtain a measure of this irreversible energy loss associated with all resistive electric circuits.

Consider again the circuit containing a resistor R and a capacitor C connected in series across a battery of emf $\mathscr{E}$ (Figure 23-10). To obtain a measure of the Joule heating in this circuit, let us multiply the circuit equation in (23-19) throughout by the instantaneous value of the current i. The result is

$$i\mathscr{E} = \frac{qi}{C} + Ri^2 \tag{23-22}$$

where use has been made of (23-1). We shall now establish that the term Ri^2 represents the rate at which energy in the form of heat is dissipated in the resistor by relating the other two terms $\mathscr{E}i$ and qi/C to the energy associated with the battery and the capacitor, respectively.

According to the definition in Section 23-2, the emf $\mathscr{E}$ of a battery represents the work per unit charge performed by the battery in carrying a charged particle between its terminals. The quantity $q\mathscr{E}$ thus represents the work that the battery carries out to transport charge q so that its rate of doing work, or power, P_B, is

$$P_B = \frac{d}{dt}(\mathscr{E}q) = \mathscr{E}i \tag{23-23}$$

Comparison with (23-22) shows that the left-hand side of that relation represents, therefore, the rate at which the battery carries out work.

In a similar way we can show that the term qi/C in (23-22) is the rate at which the energy stored in the capacitor is increasing. According to (22-9), the energy U stored in a capacitor is $q^2/2C$, and hence the rate of increase of U is

$$\frac{dU}{dt} = \frac{d}{dt}\left(\frac{1}{2C}q^2\right) = \frac{q}{C}\frac{dq}{dt} = \frac{qi}{C} \tag{23-24}$$

where the last equality follows by use of (23-1). Thus the first term on the right-hand side of (23-22) is the rate at which energy is being stored in the capacitor.

Now from a physical point of view the work expended by the battery must show up either as energy stored in the capacitor or else as energy lost in the resistor by electron collisions with the lattice. But, according to (23-24), the term qi/C on the right-hand side of (23-22) represents the rate at which the external agent (the battery in the present case) carries out work on the capacitor. Therefore since $\mathscr{E}i$ represents the power output of the battery we conclude that the residual term in (23-22), namely Ri^2, *must* represent the rate at which energy is dissipated in the resistor. Thus we have established that:

The rate P_R at which energy is dissipated in the form of heat in a resistor of resistance R through which flows a current i is given by

$$P_R = Ri^2 \tag{23-25}$$

Note that the power Ri^2 dissipated in the resistor is invariably positive, regardless of the direction of the current. Thus it represents an irreversible energy loss in the thermodynamic sense.

Example 23-9 Show that if a capacitor is charged by a battery of emf $\mathscr{E}$ through a resistor R, as in Figure 23-10, the total energy lost as Joule heat is numerically equal to that stored in the capacitor.

Solution One way of establishing this property is by making use of the explicit formulas in (23-20) and (23-21) and integrating (23-23) through (23-25).

An alternate and simpler way is by integrating both sides of (23-22) for all positive times:

$$\int_0^\infty \mathscr{E} i \, dt = \int_0^\infty \frac{qi}{C} \, dt + \int_0^\infty Ri^2 \, dt \tag{23-26}$$

Now the term on the left is the total work carried out by the battery: $\mathscr{E}q(\infty) = \mathscr{E}^2 C$, where $q(\infty) = C\mathscr{E}$ is the final charge on the capacitor. Correspondingly, the first term on the right is, according to (23-24), the final energy stored in the capacitor. Since this has the value $C\mathscr{E}^2/2$, we find by substitution into (23-26) that

$$\int_0^\infty Ri^2 \, dt = C\mathscr{E}^2 - \frac{1}{2} C\mathscr{E}^2 = \frac{1}{2} C\mathscr{E}^2$$

which is the desired result. Note that this fact, that one half of the total energy output of the battery is dissipated in the resistor, is independent of the resistance R. It is a price that must be paid each time a capacitor is charged.

23-8 Temperature variation of conductivity

In connection with the discussion of Ohm's law in (23-9) we have noted that the conductivity σ of a conductor is a constant independent of the strength of the electric field $\mathbf{E}$ and of the current density $\mathbf{j}$. In terms of (23-11) this means that the resistance R of a resistor is independent of the current i and the potential difference V developed across it. A graph of the current flowing through a resistor as a function of the potential difference V must therefore have the form of a straight line, as in Figure 23-12. Materials for which the plot of current versus voltage is a straight line are said to be *linear materials.*

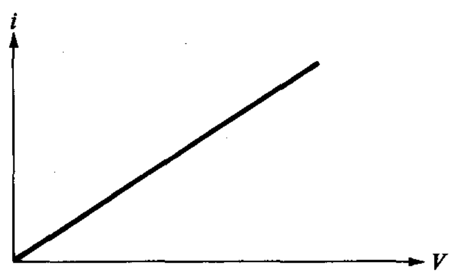


Figure 23-12

Now even though for many conductors the conductivity σ is a constant as far as current and electric field are concerned, in general σ varies with temperature. For most conductors, the conductivity rises as the temperature falls, or equivalently, the resistance R rises with temperature. Experiment shows further that if a material, which at some reference temperature T_0 has a resistivity ρ_0, undergoes a small temperature change ΔT, then the as-

sociated fractional change in resistivity, $\Delta\rho/\rho_0$, is directly proportional to ΔT. That is,

$$\Delta\rho = \alpha\rho_0\,\Delta T \tag{23-27}$$

The coefficient of proportionality α is known as the *coefficient of resistivity.* Table 23-2 lists values for α at 293 K for a number of metals. The fact that α is positive for each of the materials listed implies, according to (23-27), that the resistivity rises with temperature. For some materials (for example, amorphous carbon with $\alpha \cong -5\times 10^{-4}$/K) the coefficient of resistivity is found to be negative. This implies that it is the conductivity σ that rises with temperature for these cases. In this connection it will be established in the problems that in terms of σ (23-27) may be expressed in the equivalent form

$$\Delta\sigma = -\alpha\sigma_0\,\Delta T \tag{23-28}$$

where $\sigma_0 = 1/\rho_0$ is the value for the conductivity at the reference temperature.

Table 23-2 Coefficients of resistivity at 293 K

Substance	α (1/K)
Al	4.0×10^{-3}
Cu	4.0×10^{-3}
Ag	4.0×10^{-4}
Au	3.0×10^{-4}
Fe	5.0×10^{-3}
Pb	4.2×10^{-3}
Hg	8.9×10^{-4}
W	4.7×10^{-3}

Of considerable interest in connection with a study of the temperature variation of resistivity is a phenomenon known as *superconductivity.* In 1908 Heike Kamerlingh Onnes (1853–1926) succeeded in achieving, for the first time, liquid helium temperatures (approximately 1 K) and was thus able to measure resistivities of various substances at very low temperatures. He found that, at 4.2 K, mercury (which is a solid below 234 K) undergoes a phase transition, in that for temperatures below 4.2 K its resistivity vanishes exactly. In Figure 23-13 we plot what the results of such an experiment would look like. Above 4.2 K, solid mercury behaves as a normal conductor in that its resistance rises with temperature. However, below this critical temperature its resistance drops abruptly to zero. Subsequent experiments by Kamerlingh Onnes and others have established the fact that this property of zero resistivity, or *superconductivity* as it is called today, is shared by other substances, such as Sn, Pb, and In, which become superconductors at 3.7 K, 7.2 K, and 3.4 K, respectively. Today we know that many materials including various alloys become superconductors at sufficiently low temperatures.

It has been found that superconductors have various interesting properties.

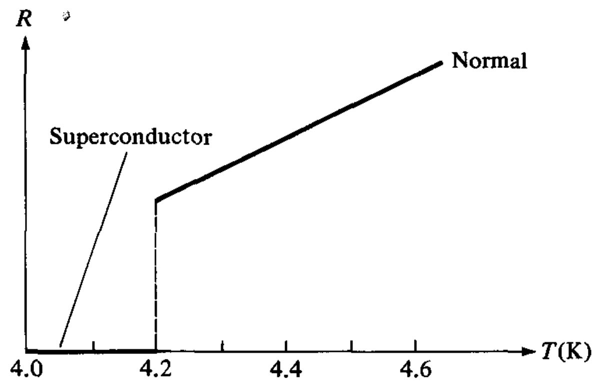


Figure 23-13

One of these is that if a current is started in a superconducting loop it will, in principle, persist forever. That is, since for a superconductor $R = 0$, it follows that the Joule heating Ri^2 must vanish exactly. Thus there is no power dissipation, and no mechanism is available to cause the current to decrease. Physically, this means that the electrons comprising the current in the superconductor undergo no energy-losing collisions with the lattice, a phenomenon that can be understood only within the framework of quantum mechanics.

23-9 Summary of important formulas

The current i flowing past a fixed point in a wire is defined by

$$i = \frac{dq}{dt} \tag{23-1}$$

and is the rate at which charge flows past this point, along a fixed and previously assigned direction. Negative values for i correspond to charge flow in the opposite sense.

The emf $\mathscr{E}$ of a battery on open circuit is related to the potential difference across its terminals by

$$\mathscr{E} = V_{AB} \tag{23-5}$$

If n represents the number of charge carriers per unit volume, each of charge $-e$, in a metal, then the current density $\mathbf{j}$ associated with their motion at the *drift velocity* $\mathbf{v}_d$ is

$$\mathbf{j} = -ne\mathbf{v}_d \tag{23-7}$$

In terms of the resistance R of a thin wire, Ohm's law is given by

$$V = Ri \tag{23-11}$$

where V is the potential difference between the ends of the wire when a current i flows through it. The power P_R dissipated in the resistor is given by

$$P_R = Ri^2 \tag{23-25}$$

QUESTIONS

1. Define or describe briefly what is meant by the following: **(a)** current; **(b)** electromotive force; **(c)** conductivity; **(d)** Ohm's law; and **(e)** Joule heating.
2. Why can the electric field inside the connecting wire in Figure 23-1b not vanish as long as the capacitor has a charge? Explain.
3. A wire is connected to one terminal of a battery, with the other end of the wire remaining free. Will a current flow in the wire? Explain.
4. Why must a circuit be closed in order for a steady current to flow in it? If a circuit is broken at some point, electric charge collects at the break. Why?
5. Explain why, in general, the electric field is *nonvanishing* in the region of space *outside* the circuit elements of a circuit in which a current flows?
6. What, in physical terms, is the distinction between the emf of a battery and the electric potential between its terminals? To be specific, consider a chemical battery.
7. Explain in physical terms why you would expect the conductivity of a conductor to be a decreasing function of temperature. Can you account for the fact that for some conductors σ is an increasing function of temperature?
8. A battery of emf $\mathscr{E}$ is connected across a wire. What happens to the current if **(a)** the length of the wire is doubled? **(b)** the diameter of the wire is doubled?
9. What is the distinction between the drift velocity of the electrons in a conductor and the velocity associated with their random motions?
10. A battery of emf $\mathscr{E}$ is connected across a wire of length l, of cross-sectional area A, and of conductivity σ. What is the effect on the drift velocity if **(a)** l is increased? **(b)** A is increased? **(c)** σ is increased?
11. In view of the fact that the Joule heating in a resistor is Ri^2, why is it that if a resistor is connected across a battery, then these losses actually *decrease* as the resistance *increases*?
12. Explain in physical terms why the initial current flow in the circuit in Figure 23-10 is the same as if the capacitor were not present. Similarly explain why the saturated value of the current is the same as if the resistor were not in the circuit.
13. Consider the circuit in Figure 23-14 and suppose that at $t = 0$ the switch is closed. Explain in physical terms why you would expect the initial value of the current to be $\mathscr{E}/R_2$.

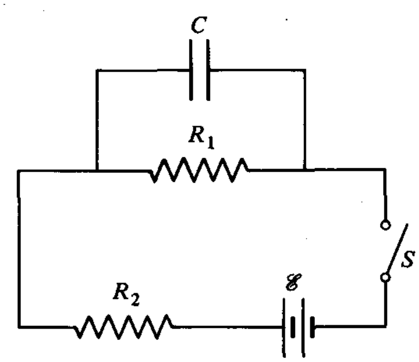


Figure 23-14

14. Explain why a battery must supply $2U$ units of energy to charge up a capacitor so that it has a final energy U. What is the maximum fraction of the output energy $2U$ of the battery that is convertible to useful work?
15. Consider the circuit in Figure 23-14. Explain, in physical terms, why you would expect after a long time that the current through the battery is $\mathscr{E}/(R_1 + R_2)$. What would be the charge on the capacitor at this time?
16. A 75-watt and a 100-watt light bulb

are designed to be powered by a 110-volt line. Through which will the higher current flow?

17. Consider the electron path $\overline{ab}$ in Figure 23-4. Explain why this path cannot be straight! Which way would its curvature be?

18. Consider an electron whose motion contributes to the current in a wire. Will this electron collide with electrons as well as with ions? Do both types of collisions contribute to the Joule heat? Explain.

19. Explain the negative sign in (23-14) and explain why this relation is *not* in contradiction with the defining relation (23-1). (*Hint*: Do the symbols q in these two relations represent the same physical quantity?)

20. If the resistance R in Figure 23-10 were doubled, explain what would happen to **(a)** the initial current through the battery; **(b)** the time constant τ; **(c)** the final charge on the capacitor; and **(d)** the energy output of the battery.

PROBLEMS

1. A battery having an emf of 100 volts is connected across a resistor of resistance $500\,\Omega$. Calculate **(a)** the current in the wire and **(b)** the current density in the wire, assuming it has a cross-sectional area of 5.0 mm^2.

2. A current of 0.5 mA flows in a wire. **(a)** How much charge passes a given point in 10 minutes? **(b)** How many electrons pass a given point in the wire in 1 second?

3. A wire of cross-sectional area $2.0 \times 10^{-5}\text{ m}^2$ is welded end-to-end to a second wire of cross-sectional area $4.0 \times 10^{-5}\text{ m}^2$. If a steady current of 5.0 amperes flows through this composite wire, what is the current density in each segment? Assume uniform flow.

4. A wire is made of copper and has a cross-sectional area of 0.2 cm^2. Use the fact that the atomic mass of copper is 64 and assume that it has a density of 9.0 g/cm^3.

(a) What is the number of copper atoms per unit volume in the sample?

(b) If a current of 10 amperes flows along this wire, what is the current density?

(c) Assuming one free electron per atom in copper, calculate the drift velocity of the electrons.

5. A 10-volt battery is connected across a 3-Ω resistor.

(a) How much current flows through the resistor?

(b) How much charge flows through the battery during a 1-minute time interval?

(c) How many electrons flow through the resistor in 15 minutes?

6. A uniform current of 5.0 amperes flows along an aluminum wire of cross-sectional area $2.0 \times 10^{-5}\text{ m}^2$.

(a) What is the current density in the wire?

(b) What is the electric field strength in the wire? Use the data in Table 23-1.

(c) If the wire has a length of 10 meters, what is the potential drop across it?

7. A silver wire and an aluminum wire have the same length and the same cross-sectional areas. What is the ratio of the currents that flow through them if they are separately connected to two batteries each with an emf of 5 volts?

8. Show that the unit of the farad-ohm (F-Ω) is the same as the second.

9. A copper wire has a resistance of $10\ \Omega$ and a length of 3 meters. It is then stretched so that its length becomes 12 meters. Assuming that the volume of the wire is *not* altered in the stretching process, calculate the final resistance of the wire. Assume that the temperature of the copper remains constant.

10. A wire has a length of 100 meters and a cross-sectional area of $10^{-6}\ \text{m}^2$. It is found that a 50-volt battery will produce a 10-ampere current when connected across the wire. **(a)** Calculate the conductivity of the wire. **(b)** What is its resistivity?

11. A certain light bulb dissipates 100 watts when connected across a 100-volt line. **(a)** What is the filament resistance? **(b)** What current flows through the filament?

12. Calculate the energy dissipated as Joule heat in 1 hr in the resistor in Problem 1.

13. A 15-volt battery is connected across two resistors, as shown in Figure 23-15.

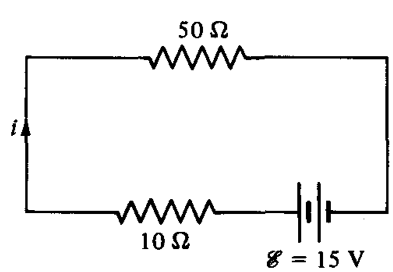


Figure 23-15

(a) In terms of the current i flowing in the circuit, what is the potential drop across the $10\text{-}\Omega$ resistor?
(b) Repeat (a) for the other resistor.
(c) Making use of (23-5) show that a current of 0.25 ampere flows in the circuit.

14. Consider again the circuit in Figure 23-15.
(a) What is the power expended in the $10\text{-}\Omega$ resistor?
(b) How much energy in the form of heat is dissipated in 1 hr in the $50\text{-}\Omega$ resistor?
(c) How much energy will be expended by the battery if it operates for 10 hr?

15. What must be the resistance of a resistor that, when connected in series with a $2.5\text{-}\mu\text{F}$ capacitor across a battery, will yield a time constant for the circuit of **(a)** 1 μs? **(b)** 1 second?

16. A $5.0\text{-}\mu\text{F}$ capacitor is charged up by a 100-volt battery and is then disconnected from the battery and reconnected across a $200\text{-}\Omega$ resistor.
(a) What is the original charge on the capacitor?
(b) What is the time constant τ for the circuit?
(c) What is the current through the resistor at $t = \tau$?

17. A charge of $10\ \mu\text{C}$ is placed on a $0.05\text{-}\mu\text{F}$ capacitor and then allowed to discharge through a $50\text{-}\Omega$ resistor.
(a) What is the time constant for this circuit?
(b) What is the initial value for the current?
(c) How much energy is originally stored in the capacitor?
(d) What is the rate at which energy is dissipated in the resistor at any time t?

18. By use of (23-17) show, by explicit calculation, that all the energy originally stored in the capacitor is dissipated as Joule heat in the resistor.

19. For the circuit in Figure 23-10, assuming $\mathscr{E} = 100$ volts, $C = 5.0\ \mu\text{F}$, and $R = 200\ \Omega$, calculate:
(a) The initial current in the circuit.
(b) The charge on the capacitor at the times 2.0×10^{-4} second, and 2.5×10^{-3} second.
(c) The current in the circuit at the times in (b).

20. Consider again the circuit of Problem 19.

(a) What is the final value of the potential across the capacitor?
(b) What is the final charge on the capacitor?
(c) What fraction of the energy put out by the battery is stored in the capacitor? What happens to the remainder of this energy?

21. Consider again the circuit in Problem 19.
(a) What is the charge on the capacitor at time t?
(b) What is the current at time t?
(c) Calculate the total energy lost in the resistor.

22. Suppose that the capacitor in Figure 23-10 has an original charge Q_0. Assume that at $t=0$ the switch is closed.
(a) Show that (23-19) is still the circuit equation.
(a) Show that

$$q(t)=C\mathscr{E}+(Q_0-C\mathscr{E})e^{-t/RC}$$

(c) Calculate the current at any time t and explain in physical terms why if $Q_0=C\mathscr{E}$, there is no current at any time.

23. Consider again the situation in Problem 22 for the special case $Q_0=2C\mathscr{E}$.
(a) Show that the current is

$$i(t)=-\frac{\mathscr{E}}{R}e^{-t/RC}$$

and discuss the physical significance of the minus sign.
(b) Calculate the total energy expended in the resistor.
(c) Show that the amount of work carried out *on* the battery is $C\mathscr{E}^2$.
(d) Compare your results of (b) and (c) to the difference between the original energy stored in the capacitor and its final energy. Account for any differences.

24. Suppose that the switch S in Figure 23-10 is closed at time $t=0$, and then reopened at time $t=T$.
(a) Show that for $t<T$ the current through the resistor is

$$i(t)=\frac{\mathscr{E}}{R}e^{-t/RC} \qquad (t<T)$$

(b) Show that at $t=T$, the ratio of the energy stored in the capacitor to that dissipated in the resistor is $\tanh(T/2RC)$.

25. Assuming that the capacitor in Figure 23-10 is a parallel-plate capacitor of area A and plate separation d, calculate the electric field in the region between the plates at any time t. Assume that the switch is closed at $t=0$.

26. A spherical capacitor of inner radius a and outer radius b is connected through a resistor R to a battery of emf $\mathscr{E}$, as shown in Figure 23-16.

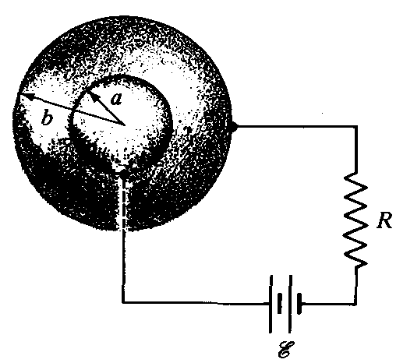


Figure 23-16

(a) Show that the time constant τ for this circuit is

$$\tau=4\pi\epsilon_0 R\frac{ab}{b-a}$$

(b) Calculate the electric field in the region between the conducting spheres at any time t. Assume that the current first begins to flow at $t=0$.

*27. Consider two capacitors C_1 and C_2 connected to a resistor R as in Figure 23-17. If i is the current directed as shown, and q_1 and q_2 are the charges on C_1 and C_2 at time t:

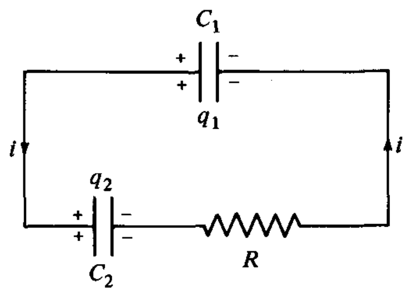


Figure 23-17

(a) Show that the circuit equation is

$$\frac{q_1}{C_1} = Ri + \frac{q_2}{C_2}$$

where if q_1, $q_2 > 0$, then the left-hand plates of C_2 and C_1 have the positive charge.

(b) Show that this may be reexpressed as

$$R\frac{dq_2}{dt} + q_2\left(\frac{1}{C_2} + \frac{1}{C_1}\right) - \frac{Q_0}{C_1} = 0$$

where Q_0 is the total charge on the two capacitors originally.

(c) Solve the circuit equation in (b) and thus find a formula for the current through R at any time t. Assume that at $t = 0$, $q_2 = 0$, so that Q_0 is the initial charge on C_1.

***28.** For the physical situation described in Problem 27, calculate in terms of Q_0, C_1, C_2, and R:

(a) The initial current in the circuit.

(b) The final charge on each capacitor.

(c) The total energy dissipated in the resistor.

29. The resistivity of aluminum at 20°C is 2.8×10^{-8} Ω-m, and its coefficient of resistivity at the same temperature is 4.0×10^{-3}/°C. Calculate the resistivity of this substance at **(a)** 0°C and **(b)** 100°C.

30. If α is the coefficient of resistivity of a substance, derive (23-28) by use of (23-27) and (23-10).

31. Show that if α is the coefficient of resistivity and β is the coefficient of linear expansion of a given wire, and if the temperature is raised by a small amount ΔT, then the fractional change in the wire's resistance, $\Delta R/R_0$, is

$$\frac{\Delta R}{R_0} = (\alpha - \beta)\,\Delta T$$

(*Note*: For typical metals the coefficient of linear expansion β has a value of about 10^{-5}/K. On the other hand, we see from Table 23-2 that $\alpha \approx 10^{-3}$/K, and thus $\alpha \gg \beta$.)

32. A copper wire at room temperature carries a current of 2.0 amperes if a 50-volt battery is connected across it.

(a) What is the resistance of the wire at room temperature?

(b) If the temperature of the wire is raised by 30 K, what is its resistance now?

(c) What current would flow (assuming again a 50-volt battery) through this wire at the elevated temperature?

33. A tungsten wire has a resistance of 100 Ω at 273 K. Calculate its resistance at 1500 K, assuming that (23-27) is valid for such a large temperature increment.

***34.** The motion of an electron in a conductor in the presence of an electric field $\mathbf{E}_0$ is often described in terms of a model originally proposed by Drude. In his model Drude assumed that a frictional force

$$\mathbf{F} = -m\omega_c\mathbf{v}_d$$

acts on the electron. In this expression m is the electron's mass, $\mathbf{v}_d$ is its drift velocity, and ω_c is a certain parameter called the *collision frequency*, which represents the number of times per second that an electron collides with a lattice ion.

(a) Using the fact that an electron moves at the uniform velocity $\mathbf{v}_d$ under the combined action of this force and that due to the electric field $\mathbf{E}_0$, show that

$$\omega_c = \frac{eE_0}{mv_d}$$

(b) Show that the current density $\mathbf{j}$ in a conductor may be written in the form

$$\mathbf{j} = \frac{ne^2}{m\omega_c}\mathbf{E}_0$$

and that therefore we may express the conductivity by the formula

$$\sigma = \frac{ne^2}{m\omega_c}$$

(c) Make use of this formula for σ to calculate ω_c for a copper conductor for which $\sigma = 5.9 \times 10^7\ (\Omega\text{-m})^{-1}$. Use the data for copper in Problem 4.

24 Elementary circuit theory

Die Grösse des Stromes in einer galvanischen Kette, ist der Summe aller spannungen direct, und der ganzen reducirten Lange der Kette umgekehrt proportional....

GEORG S. OHM (1787-1854)

24-1 Introduction

In Chapter 23 we studied the flow of electric currents in circuits consisting of only a single loop. These circuits, as exemplified in Figures 23-7 and 23-10, are characterized by the fact that at any given instant the same current i flows through each element of the circuit. The purpose of this chapter is to generalize these results to the case of multiloop circuits, such as that in Figure 24-1, for which the currents in the various elements are, in general, different.

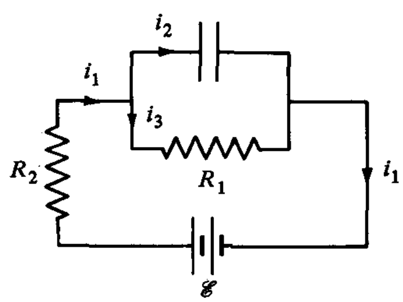


Figure 24-1

24-2 Resistors in series and parallel

To open the discussion of multiloop circuits, let us consider first the problem of replacing certain combinations of resistors by an equivalent single resistor.

Two resistors R_1 and R_2 can be connected to each other in either of two ways. If they are connected as in Figure 24-2a they are said to be in *series*, whereas if they are connected as in 24-2b they are said to be connected in *parallel*. The current through each of two resistors in series must be the same, whereas the current in each of the two resistors connected in parallel will, in general, be different. On the other hand, the potential drops across two resistors in parallel must be the same since each end of each resistor is at the same potential as is that end of the other with which it is in electrical contact. By contrast, the potential drops across two resistor in series are, in general, different.

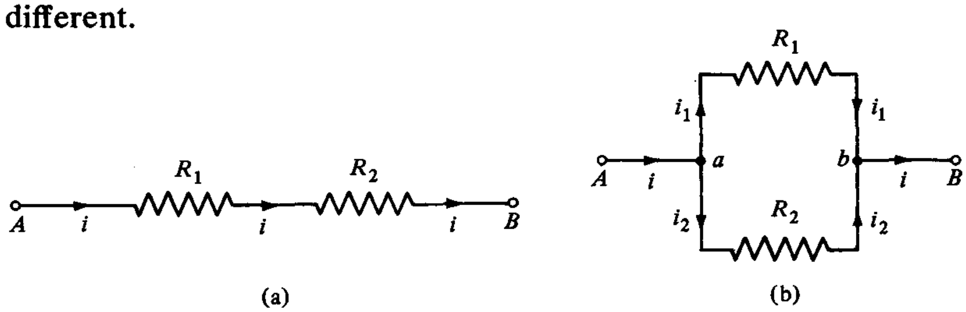


Figure 24-2

We shall now establish that if two resistors R_1 and R_2 are connected in series, then they can be replaced by a single equivalent resistor with resistance

$$R = R_1 + R_2 \qquad \text{(series connection)} \tag{24-1}$$

whereas if they are in parallel, the equivalent resistance R is

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \qquad \text{(parallel connection)} \tag{24-2}$$

Note that the rules for combining resistors in series and parallel are the reverse of those governing capacitors, as given in (22-6) and (22-7). Table 24-1 summarizes these rules for both resistors and capacitors.

Table 24-1

Connection	*Equivalent capacitance C*	*Equivalent resistance R*
Series	$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$	$R = R_1 + R_2$
Parallel	$C = C_1 + C_2$	$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$

To derive (24-1) suppose that the endpoints A and B of the resistors in series in Figure 24-2a are kept at a fixed potential difference V_{AB}. The current flow i must then have a magnitude such that the *sum* of the potential drops iR_1 and iR_2 across the two resistors has the value V_{AB}. Hence,

$$i = \frac{V_{AB}}{R_1 + R_2}$$

If now R_1 and R_2 are replaced by a single resistor R, the current will adjust to some new value V_{AB}/R. For the equivalent resistor R, by definition, the current must assume the above value $V_{AB}/(R_1 + R_2)$. Hence (24-1) follows.

The derivation of (24-2) proceeds along similar lines. Suppose that points A and B in Figure 24-2b are kept at some fixed potential difference V_{AB}. The current i that leaves point A will travel to point a, where it branches into two currents: one of strength i_1 going through R_1 and the second of strength i_2 going through R_2. At point b these currents merge again and continue on to point B. If steady-state conditions are assumed to prevail, then there will be no accumulation of charge at any point in the circuit, including points a and b. Hence it follows that

$$i = i_1 + i_2 \tag{24-3}$$

which is another way of saying that all of the charge that arrives at the point a via the current i leaves this point by means of either the current i_1 or the current i_2. Since the potential drops across a resistor R is iR, it follows by reference to the figure that

$$V_{AB} = i_1 R_1 = i_2 R_2 = iR \tag{24-4}$$

where R represents the resistance of the single resistor which, when connected across AB, would draw the current i. Finally, combining (24-3) and (24-4), we conclude that the resistance R equivalent to R_1 and R_2 in parallel is given by (24-2).

By repeated application of the results in (24-1) and (24-2), it is also possible to calculate the equivalent resistance associated with combinations of resistors connected in certain ways. Thus, the equivalent resistance R_a for R_1, R_2, and R_3 connected as in Figure 24-3a is

$$R_a = R_1 + R_2 + R_3$$

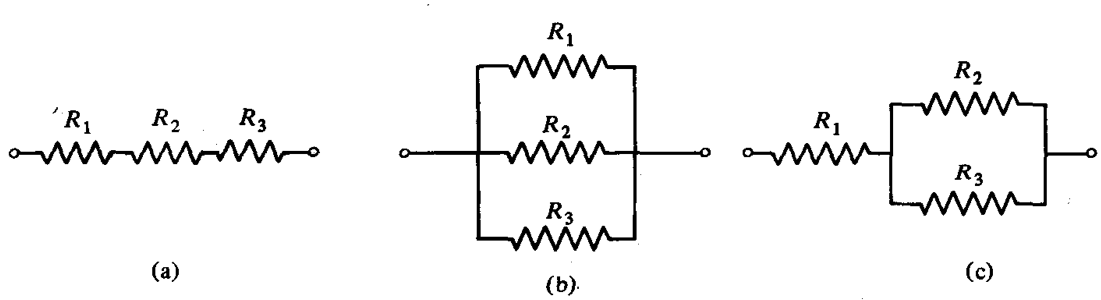


Figure 24-3

and that for the connection in Figure 24-3b satisfies

$$\frac{1}{R_b} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$$

Similarly, the equivalent resistance R_c for the resistors connected as in Figure 24-3c is

$$R_c = R_1 + \left(\frac{1}{R_2} + \frac{1}{R_3}\right)^{-1} = R_1 + \frac{R_2 R_3}{R_2 + R_3}$$

where the second term is the equivalent resistance of R_2 and R_3 in parallel.

Example 24-1 What current i is drawn from a 10-volt battery if 10 resistors, each of resistance 20 Ω, are connected across it in: (a) Series? (b) Parallel?

Solution

(a) For resistors in series, the equivalent resistance, according to (24-1), is the sum of the individual resistances. Thus the equivalent resistance R is

$$\begin{aligned} R &= 20\,\Omega + 20\,\Omega + \cdots + 20\,\Omega \\ &= 200\,\Omega \end{aligned}$$

According to Ohm's law, the current drawn from the battery is

$$i = \frac{\mathscr{E}}{R} = \frac{10\text{ V}}{200\,\Omega} = 0.05\text{ A}$$

(b) For the parallel connection, the reciprocal of the equivalent resistance is given by the sum of the reciprocals of the individual resistances. Hence

$$\frac{1}{R} = \frac{1}{20\,\Omega} + \frac{1}{20\,\Omega} + \cdots + \frac{1}{20\,\Omega} = \frac{10}{20\,\Omega} = \frac{1}{2\,\Omega}$$

or, equivalently,

$$R = 2\,\Omega$$

Thus, the current i drawn by the battery is

$$i = \frac{\mathscr{E}}{R} = \frac{10\text{ V}}{2\,\Omega} = 5\text{ A}$$

and this is 100 times as large as the current that flows if the resistors are connected in series.

Example 24-2 A 100-Ω and a 300-Ω resistor are connected in parallel across a battery having an emf of 50 volts and a negligible internal resistance (see Figure 24-4).

(a) What current flows through the battery?
(b) What current flows in each resistor?
(c) Calculate the ohmic losses in each resistor.
(d) Calculate the power output of the battery.

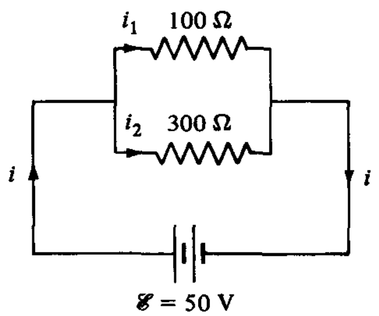


Figure 24-4

Solution

(a) According to (24-2), the resistance R equivalent to the two resistors is

$$R = \frac{R_1R_2}{R_1+R_2} = \frac{100\ \Omega \times 300\ \Omega}{100\ \Omega + 300\ \Omega} = 75\ \Omega$$

and hence the current i through the battery is

$$i = \frac{\mathscr{E}}{R} = \frac{50\ \text{V}}{75\ \Omega} = \frac{2}{3}\ \text{A}$$

(b) Since the potential drop across the two resistors is the same as the emf of the battery, it follows that

$$50\ \text{V} = (100\ \Omega) \times i_1 = (300\ \Omega) \times i_2$$

and this leads to

$$i_1 = \frac{1}{2}\ \text{A} \qquad i_2 = \frac{1}{6}\ \text{A}$$

(c) The Joule heat is Ri^2, so the power P_1 dissipated in the 100-Ω resistor is

$$P_1 = (100\ \Omega) \times i_1^2 = (100\ \Omega) \times \left(\frac{1}{2}\ \text{A}\right)^2 = 25\ \text{W}$$

while for the other it is

$$P_2 = (300\ \Omega) \times i_2^2 = (300\ \Omega) \times \left(\frac{1}{6}\ \text{A}\right)^2 = 8.3\ \text{W}$$

(d) The power output P_B of the battery is

$$P_B = \mathscr{E}i = (50\ \text{V}) \times \left(\frac{2}{3}\ \text{A}\right) = 33.3\ \text{W}$$

and this is the same as the sum $(P_1 + P_2)$ of the rates at which energy is dissipated in the resistors, as it must be.

24-3 Kirchhoff's rules

The purpose of this section is to introduce two rules, known as Kirchhoff's rules, which are indispensable for calculating the current flows in multiloop circuits.

Consider, in Figure 24-5, a three-loop network which contains two batteries of respective emf's $\mathscr{E}_1$ and $\mathscr{E}_2$ with polarities connected as shown, six resistors $R_1, R_2, \ldots, R_6$, and two additional resistors r_1 and r_2, which represent the internal resistances of the batteries. The points C, D, H, and G, where three or more wires meet, are called the *junctions* of the circuit, and the closed circuit paths $ABCHA$, $CDGHC$, and $DEFGD$ define three possible *loops* in this circuit. The circuit path connecting any two neighboring junctions is called a *branch* of the circuit, and thus, for example, CH and $DEFG$ are two branches.

Since the current that flows in a given branch will in general differ from that flowing in any other, in order to analyze this circuit it is necessary first to assign a mathematical symbol and a sense of direction for the current in each branch. (Should the assignment of this direction for any current be incorrect, the subsequent calculations would show that current to have a negative value.) Since there are altogether six branches in the network in Figure 24-5, there will be six currents; these have been labeled i_1, i_2, i_3, i_4, i_5, and i_6 and have been arbitrarily assigned the directions shown. The problem now is to evaluate these six currents in terms of the various resistances and battery emf's.

The procedure that is most commonly used to obtain these currents involves the application, to the given network, of two principles which are known as *Kirchhoff's rules*. Generally speaking, these rules, when properly applied, lead to just the right number of linearly independent relations to determine these currents. In particular, for the network in Figure 24-5, this means, as will be seen below, that Kirchhoff's rules yield six linearly independent relations for the six unknown currents $i_1, i_2, \ldots, i_6$.

The first of these two rules is based on the physical fact that under steady conditions, electric charge does not accumulate at any point in a circuit except possibly on the plates of a capacitor. To satisfy this condition at the junctions of the circuit, it is necessary that the net current flowing into each junction must vanish. Hence follows Kirchhoff's rule 1:

At any junction in a network, the sum of the currents approaching this junction must be equal to the corresponding sum of the currents leaving the same junction.

By way of illustration, consider the circuit in Figure 24-5. Applying this rule successively to the junctions C, D, G, and H, we find that the six unknown currents must satisfy the four relations:

$$\begin{aligned} i_1 &= i_2 + i_3 \\ i_2 + i_5 + i_6 &= 0 \\ i_4 &= i_5 + i_6 \\ i_3 &= i_1 + i_4 \end{aligned} \tag{24-5}$$

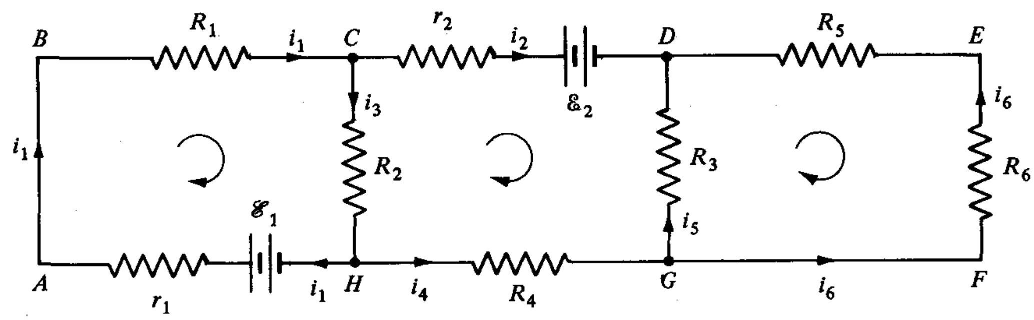


Figure 24-5

In order to obtain additional relations among these currents it is necessary to utilize the second rule. Imagine an external agent taking a unit (positive) charge around a given loop of a circuit in (say) a clockwise sense. Since the electrostatic field is conservative, it follows that the work he must carry out against the electrostatic field in traversing the loop vanishes. But on going through a resistor R carrying a current i, the work that he carries out will be $-Ri$ for traversals along the direction of the current and $+Ri$ for traversals in the opposite direction. Similarly, according to the analysis of Section 23-2, on going *through* a battery of emf $\mathscr{E}$ the work he must carry out will be $+\mathscr{E}$ if the battery is traversed from the negative (−) to the positive (+) terminals and $-\mathscr{E}$ if it is traversed in the opposite direction. Hence follows Kirchhoff's rule 2:

As a loop is traversed in (say) a clockwise sense the algebraic sum of the potential drops across all batteries in the loop plus the algebraic sum of the potential drops across all resistors in the loop will vanish provided the signs of the potential drops across the resistors and the seats of emf are as defined immediately above.

As an illustration, if the single-loop circuit in Figure 23-6 is traversed in a clockwise sense, this rule leads directly to the known result $-Ri + \mathscr{E} = 0$. If the loop is traversed counterclockwise against the direction of the current, we obtain the equivalent result: $Ri - \mathscr{E} = 0$.

Let us return now to the circuit in Figure 24-5. The application of rule 2 successively to the loops *ABCHA*, *CDGHC*, and *DEFGD* (assuming a clockwise traversal in each case) leads to

$$\begin{aligned} \mathscr{E}_1 - i_1R_1 - i_3R_2 - i_1r_1 &= 0 \\ -\mathscr{E}_2 + i_5R_3 + i_4R_4 + i_3R_2 - i_2r_2 &= 0 \\ i_6R_5 + i_6R_6 - i_5R_3 &= 0 \end{aligned} \tag{24-6}$$

The fact that these relations, when combined with (24-5), comprise seven relations to determine six currents is no essential problem, since they are *not* linearly independent. For the addition of the second of (24-5) to the third

yields the relation $i_2 = -i_4$, which when substituted into the first yields the fourth. Hence the systems of equations in (24-5) and (24-6) are consistent and lend themselves, in principle, to a unique solution for the six unknown currents in terms of the battery emf's $\mathscr{E}_1$ and $\mathscr{E}_2$ and the values for the eight resistors. The actual solution for this circuit is somewhat complex and it is not particularly instructive to pursue the matter here in any further detail.

24-4 Applications of Kirchhoff's rules

In this section the general utility of Kirchhoff's rules will be illustrated by reference to a number of simple multiloop networks.

Example 24-3 Two resistors R_1 and R_2 are connected in parallel across a battery of emf $\mathscr{E}$ and internal resistance r. Calculate the current in each resistor.

Solution Rather than solve this problem as we did in Section 24-2 by calculating the resistance of the equivalent resistor, let us apply Kirchhoff's laws directly.

As shown in Figure 24-6, the circuit has three branches, two junctions, and two independent loops. If we assign currents i_1, i_2, and i_3 directed as shown, then an application of rule 1 to the left junction leads to

$$i_1 = i_2 + i_3$$

and an identical relation is obtained at the right-hand junction. Applying rule 2 to the upper loop we obtain

$$-R_1 i_3 + R_2 i_2 = 0$$

and an application of this rule to the lower loop yields

$$-i_1 r - i_2 R_2 + \mathscr{E} = 0$$

Solving these relations we find that

$$i_1 = \frac{\mathscr{E}}{r + \dfrac{R_1 R_2}{R_1 + R_2}}$$

$$i_2 = \frac{\mathscr{E} R_1}{r(R_1 + R_2) + R_1 R_2}$$

$$i_3 = \frac{\mathscr{E} R_2}{r(R_1 + R_2) + R_1 R_2}$$

Note that, as expected, the current i_1 through the battery is that which would flow through the resistor R equivalent to r connected in series with the combination of R_1 and R_2 in parallel. Note also that the ratio i_2/i_3 of the currents through R_2 and R_1 is equal to the ratio, R_1/R_2.

Example 24-4 A 5-volt battery and a 20-volt battery are connected as in Figure 24-7. Calculate the current through each battery and the power output of the smaller battery.

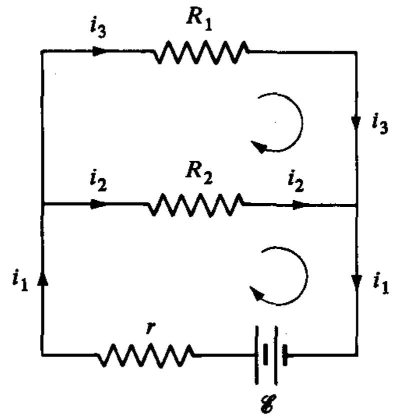


Figure 24-6

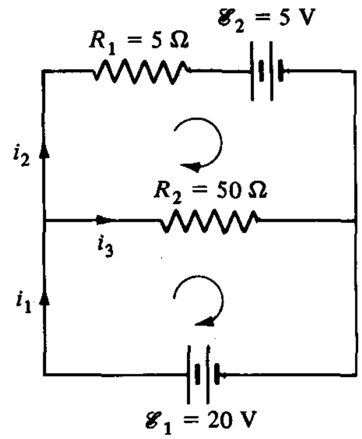


Figure 24-7

Solution On assigning the currents i_1, i_2, and i_3 as shown, we find by use of Kirchhoff's first rule that

$$i_1 = i_2 + i_3$$

The application of the second rule to the upper and lower loops, respectively, leads to

$$-\mathscr{E}_2 + R_2 i_3 - R_1 i_2 = 0$$
$$\mathscr{E}_1 - R_2 i_3 = 0$$

Solving these three relations for i_1, i_2, and i_3, and inserting the given values for R_1, R_2, $\mathscr{E}_1$, and $\mathscr{E}_2$, we obtain

$$i_1 = 3.4\ \mathrm{A} \qquad i_2 = 3.0\ \mathrm{A} \qquad i_3 = 0.40\ \mathrm{A}$$

Since the current i_2 is positive, it follows by reference to the figure that this current is directed (through the smaller, or 5-volt, battery) from the positive to the negative terminal. This means that its power output P_2 is negative and has the value

$$P_2 = -i_2\mathscr{E}_2 = -(3.0\ \mathrm{A}) \times (5.0\ \mathrm{V}) = -15\ \mathrm{W}$$

In other words, work at the rate of 15 watts is being carried out *on* this battery. The power for this charging up comes from the other battery, whose power output P_1 is

$$P_1 = i_1\mathscr{E}_1 = (3.4\ \mathrm{A}) \times (20\ \mathrm{V}) = 68\ \mathrm{W}$$

Of the residual 53 watts ($= 68\ \mathrm{W} - 15\ \mathrm{W}$), the amount

$$R_1 i_2^2 = (5\ \Omega) \times (3.0\ \mathrm{A})^2 = 45\ \mathrm{W}$$

is dissipated as heat in R_1 and the amount

$$R_2 i_3^2 = (50\ \Omega) \times (0.4\ \mathrm{A})^2 = 8.0\ \mathrm{W}$$

is dissipated in R_2. Overall, energy is conserved, as it must be.

Example 24-5 Consider the "Wheatstone bridge" circuit in Figure 24-8, with the resistor R being variable. Calculate the value of R so that the current through the 100-Ω resistor is zero. Also calculate the power output of the battery.

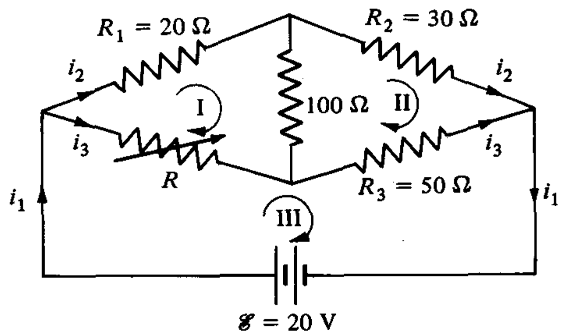


Figure 24-8

Solution Since, by hypothesis, there is no current through the 100-Ω resistor, it follows that this three-loop circuit may be described uniquely in terms of only three currents: i_1, i_2, and i_3. According to Kirchhoff's first rule, these are related by

$$i_1 = i_2 + i_3 \tag{24-7}$$

Applying Kirchhoff's second rule to loop I we find that

$$-R_1 i_2 + i_3 R = 0 \tag{24-8}$$

and, similarly, for loop II we obtain

$$-R_2 i_2 + R_3 i_3 = 0 \tag{24-9}$$

In order for (24-8) and (24-9) to be consistent, it is necessary that

$$\frac{i_2}{i_3} = \frac{R}{R_1} = \frac{R_3}{R_2} \tag{24-10}$$

or, in other words, R must satisfy

$$R = \frac{R_1 R_3}{R_2} = \frac{20\ \Omega \times 50\ \Omega}{30\ \Omega} = 33\ \Omega$$

To calculate the current i_1 let us apply Kirchhoff's second rule to loop III. The result is

$$-i_3(R + R_3) + \mathscr{E} = 0$$

Using the above value, $R = 33\ \Omega$, we thus obtain for i_3

$$i_3 = \frac{\mathscr{E}}{(R + R_3)} = \frac{20\text{ V}}{83\ \Omega} = 0.24\text{ A}$$

which, when substituted into (24-9), leads to

$$i_2 = i_3 \frac{R_3}{R_2} = (0.24\text{ A}) \times \frac{50\ \Omega}{30\ \Omega} = 0.40\text{ A}$$

Finally, substituting these values for i_2 and i_3 into (24-7), we obtain for the battery current i_1

$$i_1 = i_2 + i_3 = 0.40\text{ A} + 0.24\text{ A} = 0.64\text{ A}$$

Hence the power output P_B of the battery is

$$P_B = \mathscr{E} i_1 = (20\ \text{V}) \times 0.64\ \text{A} = 13\ \text{W}$$

When used in practice, the 100-Ω resistor in the Wheatstone bridge circuit in Figure 24-8 is usually replaced by a current-measuring device, such as a galvanometer (see Section 24-7). For given resistors R_1 and R_2 and an unknown resistor R_3, the variable resistor R is adjusted until no current flows through the galvanometer. The resistor R_3 is then determined by (24-10) to be the ratio RR_2/R_1.

24-5 Circuits with capacitors

To extend the ideas of Section 24-3 to circuits involving capacitors, let us consider the particular circuit in Figure 24-9. It has three branches; let the currents in these be represented by i_1, i_2, and i_3, directed as shown in the figure. Applying Kirchhoff's first rule to the junctions, we obtain

$$i_1 = i_2 + i_3 \tag{24-11}$$

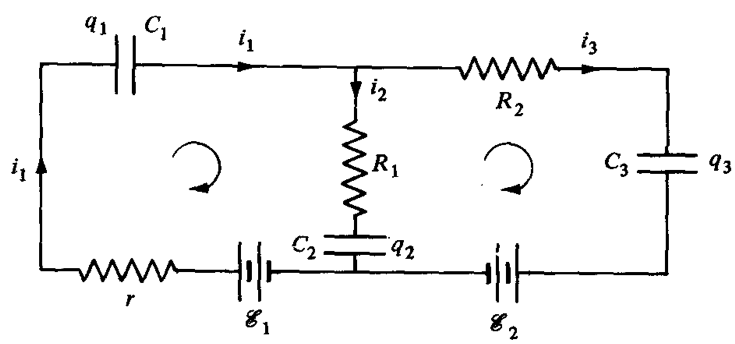


Figure 24-9

In order to apply rule 2 let us adopt the convention that the charges on the left-hand plate of C_1 and on the top plates of C_2 and C_3 are the positive charges q_1, q_2, and q_3 on these three capacitors, respectively. Then according to the definition of current,

$$i_1 = \frac{dq_1}{dt} \qquad i_2 = \frac{dq_2}{dt} \qquad i_3 = \frac{dq_3}{dt} \tag{24-12}$$

since in each case positive current corresponds to an increase of charge on the capacitor. (If, by contrast, the charge on the bottom plate of C_2 had been assumed to be the "charge" on this capacitor, then the second of these relations would have been $i_2 = -dq_2/dt$.)

Let us now generalize rule 2 to include the potential drops across capacitors and apply it to each of the two loops in Figure 24-9. Traversing the loops in a clockwise sense, for the left-hand loop we obtain

$$-i_1 r - \frac{q_1}{C_1} - i_2 R_1 - \frac{q_2}{C_2} + \mathscr{E}_1 = 0 \tag{24-13}$$

and, correspondingly, for the other loop we obtain

$$-i_3R_2 - \frac{q_3}{C_3} + \frac{q_2}{C_2} + i_2R_1 - \mathscr{E}_2 = 0 \tag{24-14}$$

In writing down these relations we have used the fact that the work required of an external agent to take a unit (positive) charge from the positive to the negative plate of a capacitor is $-q/C$ and the above convention that the top plates of C_2 and C_3 and the left plate of C_1 carry the positive charge.

The six relations in (24-11) through (24-14) constitute a set of six coupled differential equations for the three capacitor charges q_1, q_2, and q_3, and for the three currents i_1, i_2, and i_3. Their solution constitutes complete information of the charges and currents in the circuit at any time t.

24-6 The charging of a shunted capacitor

As an application of the method outlined in Section 24-5 let us consider the problem of the charging of a capacitor, whose two plates are connected by a resistor. The circuit diagram is shown in Figure 24-10. It has been assumed that the battery has a certain internal resistance r and that the resistances of the various leads can be effectively included in this resistance.

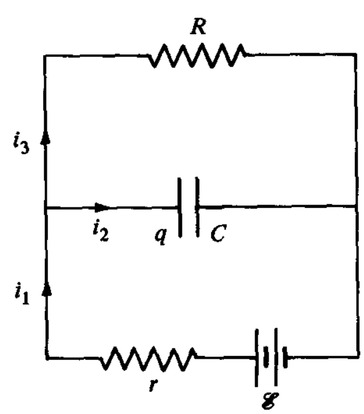


Figure 24-10

Examination of this circuit diagram shows that it has three branches, and that the currents defined there must satisfy

$$i_1 = i_2 + i_3 \tag{24-15}$$

Further, the current i_2 will be related to the charge q on the capacitor by

$$i_2 = \frac{dq}{dt} \tag{24-16}$$

provided that the left-hand plate of the capacitor is assumed to carry the positive charge. Applying rule 2 in a clockwise sense to the upper loop we

find that

$$-i_3R + \frac{q}{C} = 0 \tag{24-17}$$

and, correspondingly, for the lower loop

$$-i_1r - \frac{q}{C} + \mathscr{E} = 0 \tag{24-18}$$

The four relations, (24-15) through (24-18), constitute a complete description for this system.

Fortunately, the circuit in Figure 24-10 is sufficiently simple that it can be solved in its entirety. To this end, let us substitute the forms for i_2 and i_3 given by (24-16) and (24-17), respectively, into (24-15). Combining the resultant formula for i_1 with (24-18), we find that

$$\frac{dq}{dt} + \frac{q}{C}\left[\frac{1}{r} + \frac{1}{R}\right] = \frac{\mathscr{E}}{r}$$

To simplify matters, let us make use of the fact that in most cases of physical interest the internal resistance r is small compared to R so that the term $1/R$ is negligible in comparison with $1/r$. With this approximation this relation takes on the simpler form

$$\frac{dq}{dt} + \frac{q}{rC} = \frac{\mathscr{E}}{r} \tag{24-19}$$

Now a comparison of (24-19) with (23-19) shows that mathematically they are essentially the same. Therefore, assuming that the charge on the capacitor is initially zero—that is, $q(0) = 0$—we find the solution of (24-19) to be

$$q(t) = C\mathscr{E}(1 - e^{-t/rC}) \tag{24-20}$$

In words, this states that the charge on the capacitor approaches its asymptotic value $C\mathscr{E}$ exponentially in a time interval measured by the time constant $\tau = rC$. The substitution of (24-20) into (24-15) through (24-18) yields, for the various currents, the explicit formulas:

$$\begin{aligned}
i_1 &= \frac{\mathscr{E}}{R} + \mathscr{E}\left[\frac{1}{r} - \frac{1}{R}\right] e^{-t/rC} \\
&\approx \frac{\mathscr{E}}{R} + \frac{\mathscr{E}}{r} e^{-t/rC} \\
i_2 &= \frac{\mathscr{E}}{r} e^{-t/rC} \\
i_3 &= \frac{\mathscr{E}}{R}(1 - e^{-t/rC})
\end{aligned} \tag{24-21}$$

where the approximate formula for i_1 follows since $r \ll R$. Figure 24-11 shows a plot of the complete solution in (24-20) and (24-21). For the sake of

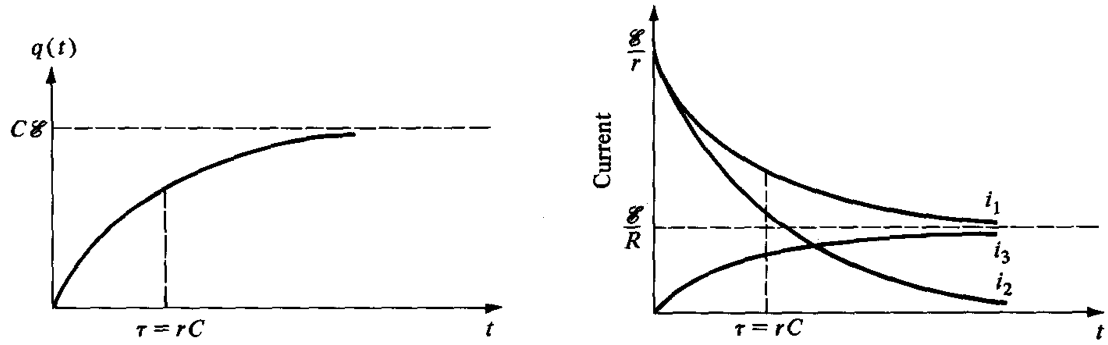


Figure 24-11

pictorial clarity only, the value $R/r = 3$ has been assumed in these graphs. Consistent with what we might anticipate on physical grounds, these graphs show that for times $t \gg rC$ the charge on the capacitor achieves its final charge $C\mathscr{E}$ and that the current i_1 through the battery becomes identical to that flowing through the resistor R. In other words, for times much greater than a time constant $\tau = rC$ the charge on the capacitor becomes the same as if the resistor R were not in the circuit, and the current through the battery assumes the value it would have if the capacitor were shorted out.

Example 24-6 For the circuit in Figure 24-10, if $\mathscr{E} = 50$ volts, $r = 1.0\ \Omega$, $R = 100\ \Omega$, and $C = 3.0\ \mu\text{F}$, calculate:

(a) The time constant of the circuit.
(b) The initial current through the battery and the resistor R.
(c) The final charge on the capacitor.

Solution

(a) The time constant τ of this circuit is rC. Hence

$$\tau = rC = (1.0\ \Omega) \times (3.0 \times 10^{-6}\ \text{F}) = 3.0\ \mu\text{s}$$

(b) The initial current through the battery may be obtained from the first of (24-21) by setting $t = 0$. Hence

$$i_1(0) = \frac{\mathscr{E}}{R} + \mathscr{E}\left(\frac{1}{r} - \frac{1}{R}\right) e^{-0} = \frac{\mathscr{E}}{r}$$

$$= \frac{50\ \text{V}}{1.0\ \Omega} = 50\ \text{A}$$

and this very large initial surge of current is the same as if neither R nor the capacitor were in the circuit. The initial current i_3 through R vanishes since the capacitor effectively shorts it out.

(c) The final charge q_f on the capacitor is

$$q_f = C\mathscr{E} = (3.0 \times 10^{-6}\ \text{F}) \times (50\ \text{V})$$
$$= 1.5 \times 10^{-4}\ \text{C}$$

Example 24-7 Consider the circuit in Figure 24-12 and suppose that at $t = 0$ the switch S is closed. Without writing down the circuit equations, calculate the following:

(a) The initial current through the battery.
(b) The steady current through the battery after a long time.
(c) The final charge on the capacitor.

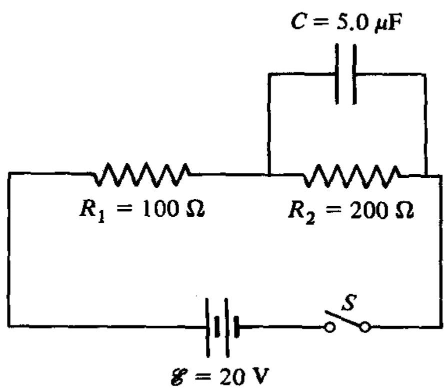


Figure 24-12

Solution

(a) Immediately after the switch is closed, the current through the battery is the same as if the capacitor were replaced by a *short*; that is, by a conductor of negligible resistance. Thus the initial current assumes the value it would have if only the 100-Ω resistor were connected across the battery. Since the total resistance across the battery is then 100 Ω, it follows that the initial current i_0 through it is

$$i_0 = \frac{\mathscr{E}}{R_1} = \frac{20\text{ V}}{100\ \Omega} = 0.2\text{ A}$$

(b) After a sufficiently long time has elapsed, the capacitor will be fully charged, and current will cease to flow in that branch of the circuit which has the capacitor. It follows that the sum of the potential drops across the 100-Ω and the 200-Ω resistors will approach the emf of the battery. Hence the steady-state current i_s through the battery is

$$i_s = \frac{\mathscr{E}}{R_1 + R_2} = \frac{20\text{ V}}{100\ \Omega + 200\ \Omega} = 0.067\text{ A}$$

(c) Reference to the figure shows that the final potential across the capacitor is the same as the final value for the iR drop across the 200-Ω resistor. Since the latter has the value

$$R_2 i_s = 200\ \Omega \times 0.067\text{ A} = 13\text{ V}$$

it follows that the final charge q_f on the capacitor is

$$\begin{aligned} q_f &= C(R_2 i_s) = 5.0 \times 10^{-6}\text{ F} \times 13\text{ V} \\ &= 6.5 \times 10^{-5}\text{ C} \end{aligned}$$

It is interesting to note that in order to obtain the initial and final values of charges and currents in a circuit, it is *not* always necessary to write down the full set of circuit equations. As illustrated in this example, by using the fact that a capacitor

acts as a short immediately after the switch is closed and as an infinite resistor after a sufficiently long time has elapsed, we can obtain a considerable amount of information about the various currents and charges in a circuit.

Example 24-8 Consider the three-loop network in Figure 24-13. Assuming that at $t = 0$ the switch S is closed, calculate:

(a) The initial current i_0 through the battery.

(b) The final charges q_1 and q_2 on the capacitors C_1 and C_2, respectively, and the final current i_f through the battery.

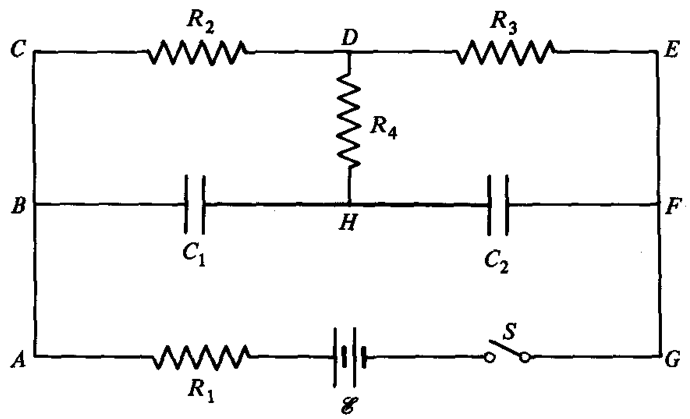


Figure 24-13

Solution

(a) Since initially there is no charge on the capacitors, it follows that the current path BHF is effectively a short. Hence, just as the switch is closed, the current will be nonzero only along the path $ABHFGA$. The total resistance along this path is R_1 and thus

$$i_0 = \frac{\mathscr{E}}{R_1}$$

must be the initial current in the battery.

(b) After the capacitors are both fully charged, no current can flow through the two branches BH and HF. Thus, after a long time a certain current i_f will flow along the path $ABCDEFGA$. Since the total resistance along this path is $(R_1 + R_2 + R_3)$, it follows that

$$i_f = \frac{\mathscr{E}}{R_1 + R_2 + R_3}$$

Since the potential drops across R_2 and R_3 must be in the same, respectively, as those across C_1 and C_2, it follows that

$$q_1 = C_1(i_f R_2) = \frac{C_1 R_2 \mathscr{E}}{R_1 + R_2 + R_3}$$

$$q_2 = C_2(i_f R_3) = \frac{C_2 R_3 \mathscr{E}}{R_1 + R_2 + R_3}$$

where in both cases the second equality is obtained by substituting the above value for i_f.

24-7 Galvanometers, ammeters, and voltmeters

In carrying out experiments on electric circuits in the laboratory, it is frequently useful to have available a means for measuring the current flow through—and the potential drop across—a given circuit element. The two instruments that have been developed for these purposes are the *ammeter* and the *voltmeter*.

An *ammeter*, as the name implies, is a device used to measure current. Normally, it is connected in series with the other circuit elements in that branch of the circuit whose current is to be measured. Correspondingly, a *voltmeter* measures the potential difference between any two points in a circuit. It is always connected across the two points whose potential difference is being measured. In a circuit diagram ammeters and voltmeters are customarily represented by the respective symbols

Figure 24-14 shows a branch of a circuit on which measurements are being made. Note that the ammeter is connected in series with the other elements in that branch through which the current flow is being measured. Its position relative to the other elements in that branch is of no significance since the same current flows through each. By contrast, the voltmeter is connected in parallel with (or across) the two points whose potential difference it is desired to measure. Thus the voltmeter in Figure 24-14 measures the potential drop across R_2. The ratio of the voltmeter reading to the ammeter reading in this circuit is numerically equal to the resistance R_2.

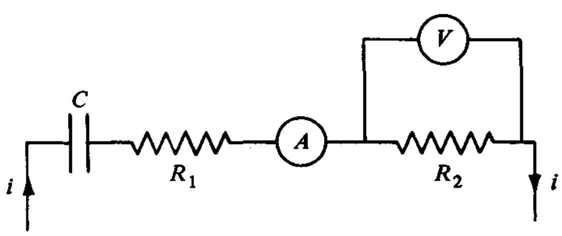


Figure 24-14

In connecting an ammeter or a voltmeter into a circuit, care must be taken so that the direction of the current is appropriate for the given instrument. To assist the experimenter, the two terminals of ammeters and voltmeters are often distinguished by a plus (+) sign on one terminal and a minus (−) sign on the other. When properly connected, current flows from the plus to the minus terminal *inside* the instrument. Also, care must be exercised in using voltmeters and ammeters so that the strength of the current does not exceed the limiting value for which the given instruments are designed.

Underlying the operation of both ammeters and voltmeters is an instru-

ment known as a *galvanometer*. The conventional symbol for a galvanometer in a circuit diagram is

Like the ammeter and the voltmeter, the galvanometer also has two terminals, which, as a rule, are also marked plus (+) and minus (−). When a current passes through a galvanometer (from the plus to the minus terminals), the pointer on the galvanometer is deflected by an amount proportional to the strength of this current. Thus, once a galvanometer has been appropriately calibrated, the strength of the current flowing through it is given by the position of the pointer on its scale. The physics underlying the operation of a galvanometer is based on the experimental fact that if a current flows through a coil of wire that is in a magnetic field, a torque proportional to the strength of this current is exerted on the coil. A detailed consideration of the principles underlying the operation of the galvanometer will be found in Chapter 26. For the moment let us view the galvanometer pragmatically as simply a device that is capable of giving a pointer reading proportional to the strength of the current flowing through it.

Consider a resistor R connected across a battery of emf $\mathscr{E}$ and of negligible internal resistance. According to Ohm's law, its resistance is the ratio of $\mathscr{E}$ to the battery current i. This suggests that to measure R we connect two galvanometers into this circuit, as shown in Figure 24-15. Applying Kirchhoff's second rule to the upper loop in the figure, we find that

$$R = \left(\frac{i_G}{i_R} - 1\right)R_G \tag{24-22}$$

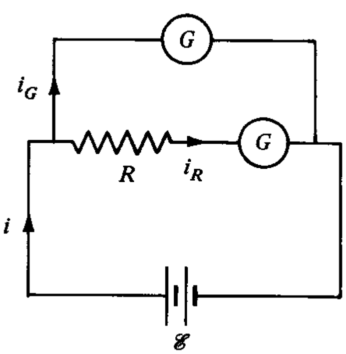


Figure 24-15

where i_G is the measured value of the current flowing through the upper galvanometer, i_R is the measured current flow through the resistor R, and R_G is the presumed known internal resistance of each of the two galvanometers. In effect, then, we can measure in this way the ratio of $\mathscr{E}/i$ for the circuit in Figure 24-15. However, this method does not yield the value of $\mathscr{E}$ or i separately. We shall now describe how by judicious use of galvanometers and resistors it is possible to construct ammeters and voltmeters and thereby to measure currents and voltages directly.

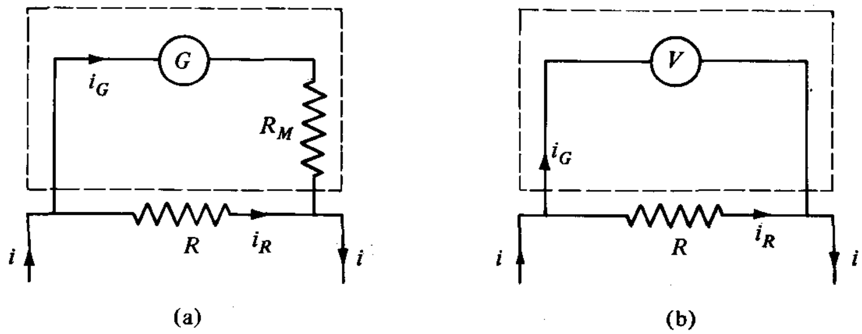


Figure 24-16

As shown in Figure 24-16a a voltmeter consists of a galvanometer connected in series with a very high resistance R_M, called the *multiplier*. Figure 24-16b shows how the circuit diagram would normally be drawn. In order not to perturb the original current that flows through R, it is necessary that R_M be selected to be very large in comparison with all other resistors in the circuit. With this choice, the potential drop across the resistance R will be essentially the same whether or not the voltmeter is present in the circuit. Applying Kirchhoff's second rule to the upper loop in Figure 24-16a we conclude that

$$i_G(R_M + R_G) = i_R R = V_R \tag{24-23}$$

with V_R the potential drop across the resistor R. Since the coefficient of proportionality $(R_M + R_G)^{-1}$ depends only on the properties of the voltmeter, it follows that the current i_G through the galvanometer is proportional to V_R. We see therefore that by suitably calibrating the voltmeter, the device in Figure 24-16 may be used to measure potential differences.

Let us now consider the operation of an ammeter. As shown in Figure 24-17a, this time a galvanometer is connected in series with the line at the same time that a very low resistor R_s, called a *shunt*, is connected in parallel with the galvanometer. Figure 24-17b shows the relevant part of the circuit diagram as it normally would be drawn. To be effective, the shunt resistance R_s must be selected to be very small compared to the resistance of the galvanometer R_G. If, in addition, $R_s \ll R$, then the original current i will not

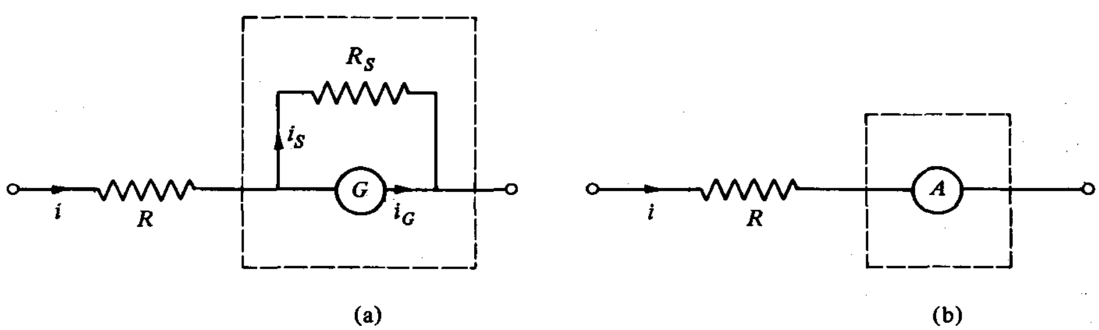


Figure 24-17

be appreciably affected by the insertion of the galvanometer into the circuit. Applying Kirchhoff's second rule to the loop in Figure 24-17a, we find that

$$i_G = \left(\frac{R_s}{R_G}\right) i_s$$

Hence the galvanometer current i_G is proportional to the shunt current i_s, with the coefficient of proportionality, R_s/R_G, depending only on the parameters characterizing the ammeter. Since i_s is, to a high degree of precision, the same as the current i flowing in the given branch of the circuit before the ammeter is inserted, it follows that a properly calibrated ammeter does indeed measure the current through that branch of the circuit into which it is inserted.

24-8 The potentiometer

It has been noted in the above discussion that the introduction of a voltmeter or an ammeter into a circuit modifies the original current flow to some extent. Even if the associated galvanometer were precisely calibrated, there would inevitably remain small experimental errors in any measurements made by their usage. The essential difficulty is that the measuring instruments themselves must draw some current from the circuit in order to operate.

By contrast, a *potentiometer* is a device that may be used to compare any two voltages without having to draw current from the original circuit in any way. If, for example, the emf of a battery is measured by use of a voltmeter, an error would arise because of the ir drop due to the internal resistance r of the battery. The corresponding measurement by use of a potentiometer would give a true reading of this emf since no current would be drawn from the battery. Evidently, the potentiometer is a very useful laboratory instrument, and the purpose of this section is to describe the principles underlying its operation.

Figure 24-18 shows the rudiments of a potentiometer. First, with the switch S open, the current i that flows around the lower loop is given, according to Ohm's law, by

$$i = \frac{\mathscr{E}}{r + R} \tag{24-24}$$

where $\mathscr{E}$ is the emf of the battery and r is its internal resistance. Suppose now that a voltage source V_1 is connected across the terminals A and B, with polarities as shown. With the switch S closed, in general, a current will flow around the upper loop and this will give rise to a nonzero reading on the galvanometer scale. Suppose that the variable contact point C is now moved along the resistor R until the galvanometer current is exactly zero. Let the symbol R_1 represent the part of the resistor R that for this condition comprises part of the upper loop. Since the current around the upper loop

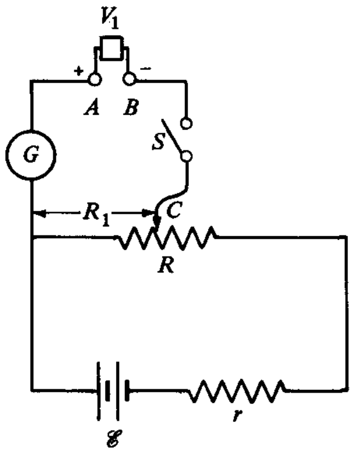


Figure 24-18

vanishes, it follows that the current i in the lower loop is precisely that in (24-24). Hence the unknown potential V_1 must be

$$V_1 = iR_1 = \frac{\mathscr{E}R_1}{R + r} \tag{24-25}$$

Note that in this process no current whatsoever is drawn from the source of the V_1 potential.

Let us now disconnect V_1 from the potentiometer and reconnect across it a second voltage source V_2. Repeating the same procedure as above, we find that in general the contact point C will be at a different position corresponding to a portion R_2 of the resistor R. The analogue of (24-25) for this second voltage source is

$$V_2 = iR_2 = \frac{R_2\mathscr{E}}{R + r}$$

Dividing this relation into (24-25) we find that

$$\frac{V_1}{V_2} = \frac{R_1}{R_2} \tag{24-26}$$

and thus the ratio of the two voltages V_1 and V_2 is independent of the battery emf $\mathscr{E}$, its internal resistance r, and the magnitude of the resistance R. It depends only on the ratio of the two intercepted resistances R_1 and R_2. In practice, a fairly accurate determination of the ratio R_1/R_2 can be obtained by making use of fixed resistors connected in series with slide wires, for which the resistance can be simply related to length. (Recall (23-12).) The ratio R_1/R_2, and thus that of V_1/V_2, can therefore be obtained with considerable precision. In this way, then, the potentiometer can be used to relate *any* potential difference to that of any convenient standard.

Figure 24-19 shows how the potentiometer can be used to calibrate accurately both an ammeter and a voltmeter. In both cases the terminals A' and B' are to be connected to the correspondingly labeled points A and B of

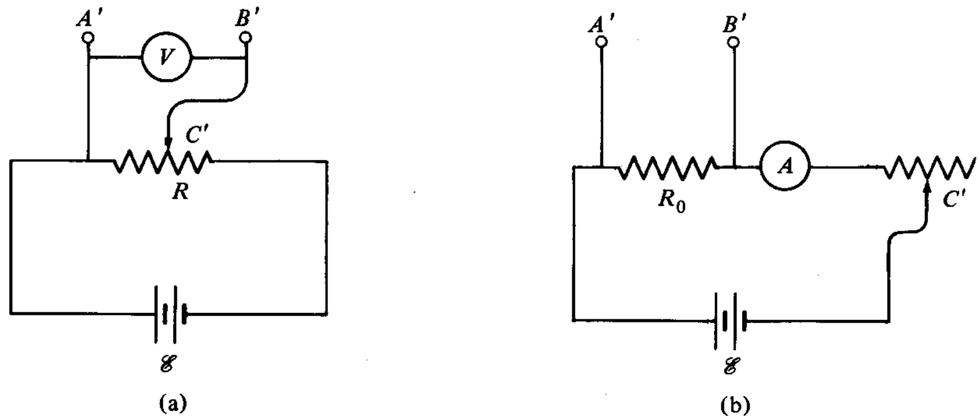


Figure 24-19

the potentiometer in Figure 24-18. As the contact point C' in Figure 24-19a is moved, the potential across the voltmeter will vary, and the voltmeter settings can be adjusted by connecting the potentiometer, for each such setting, to the points A' and B'. Similarly, in Figure 24-19b, if R_0 represents a precisely known resistor, then by moving the contact point C' both the current through the ammeter and the potential across R_0 will change. If this potential drop across R_0 is measured by use of a potentiometer, the value of the current through R_0 can thus be inferred and the ammeter calibrated.

QUESTIONS

1. Define or describe briefly the terms (**a**) junction; (**b**) resistors in series; (**c**) ammeter; (**d**) potentiometer; and (**e**) shunt resistance.
2. Two resistors of respective resistances 5 Ω and 10 Ω are connected in series. If a 2-ampere current flows through one of them, what is the current in the other?
3. If the current through each of two resistors in parallel is the same, what can be said about the two resistors? If they were connected in series, what conclusions could be drawn?
4. Twelve resistors, each of resistance R_0, are connected so that they comprise the edges of a cube. What is meant by a resistance R equivalent to this combination? How many equivalent resistances are there for this cube?
5. How much current would flow through a 10-Ω resistor that is connected in parallel with a 5-Ω resistor if a 1.0-ampere current flows through the latter?
6. Show by constructing an example that not all combinations of resistors can be decomposed into subunits of resistors in series and parallel.
7. Explain in physical terms why the current must be everywhere (except in the region between two capacitor plates) the same in a given branch of a circuit.
8. Explain in physical terms why the algebraic sum of the currents at a junction must vanish. At what other points in a circuit must this sum also vanish?
9. At what point(s), if any, in a circuit need the algebraic sum of the currents approaching this point not vanish?

10. Explain Kirchhoff's second rule in physical terms by reference to a particular two-loop circuit.
11. Would it be valid in Figure 24-5 to consider the loop *ABCDGHA* as one for which Kirchhoff's second rule would be applicable? Do you suppose if we did apply the rule to this loop that we would find a relation linearly independent of (24-6)?
12. How many loops are there altogether in the circuit in Figure 24-5? How many are there in the circuit in Figure 24-6? How many linearly independent relations do we find for each of these circuits by applying Kirchhoff's second rule?
13. A circuit contains capacitors, resistors, and batteries. Explain in physical terms why the *initial* values (after all switches are closed) for all currents are those that would be obtained if all capacitors (assumed to be originally uncharged) are shorted out—that is, if they are all replaced by resistors of zero resistance.
14. A circuit contains capacitors, resistors, and batteries. Explain in physical terms why, after the currents reach the steady state, we may calculate their values by replacing each capacitor in the original network by a resistor of infinite resistance and applying Kirchhoff's rules to the resulting circuit.
15. In a circuit a voltmeter is connected across a resistor. Assuming that the voltmeter is precisely calibrated, explain why you would expect its reading to be slightly *larger* than the potential across the resistor before the voltmeter was connected.
16. Explain in physical terms why the introduction of an ammeter into a branch of a circuit will, in general, cause a slight drop in the current in this branch. How can this error be minimized?
17. What are the comparative advantages and disadvantages of using a voltmeter or a potentiometer in measuring potential differences?
18. Explain how a potentiometer, which measures potential differences, can be used to calibrate an ammeter which measures current.
19. A voltmeter is connected across a fully charged but isolated capacitor. Describe the subsequent behavior of the voltmeter pointer. Assume that the multiplier resistance is so large that the time constant for the circuit is 2 seconds.
20. A voltmeter is connected across a resistor, and the pointer goes to maximum deflection. If the multiplier resistance is doubled, what is the new position of the pointer? What would happen if the multiplier resistance were halved?
21. If the shunt resistance in an ammeter is doubled, what happens to the pointer, assuming that it is originally at maximum deflection? What would happen if the shunt resistance were halved?
22. Why need the galvanometer in the potentiometer in Figure 24-18 *not* be calibrated precisely? Explain.

PROBLEMS

1. Find the resistance equivalent to a 3-Ω, a 5-Ω, and a 15-Ω resistor if they are connected in **(a)** series; **(b)** parallel.
2. A 10-volt battery is connected across a 10-Ω and a 40-Ω resistor connected in parallel. **(a)** What current flows through the battery? **(b)** Calculate the ratio of the currents through each resistor.

3. Three resistors, each of resistance R, are connected so that they comprise the three sides of a triangle. Calculate the equivalent resistance between any two vertices. Calculate also the ratio of the currents that would flow in each resistor if a battery were connected across any two vertices.
4. (a) Show that if R_1 and R_2 are connected in parallel, then the equivalent resistance R satisfies: $R < R_1$ and $R < R_2$.
   (b) State and derive the analogous result if R_1 and R_2 are connected in series.
5. If R_1 and R_2 are connected across a battery of emf $\mathscr{E}$, show that the ratio of the power output of the battery for the series connection to that for the parallel connection is

$$\frac{R_1R_2}{(R_1+R_2)^2}$$

6. Three 60-watt bulbs are connected in series across a 120-volt line. (a) What is the resistance of each bulb? (b) What is the current through each bulb? (c) How much power is dissipated in each bulb?
7. Repeat Problem 6, assuming that this time the bulbs are connected in parallel across the 120-volt line.
8. Consider the circuit in Figure 24-20. Calculate:
   (a) The equivalent resistance across the 12-volt battery.

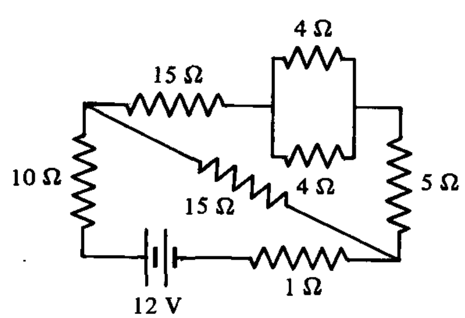


Figure 24-20

   (b) The current through the battery.
   (c) The total power dissipated in all the resistors.
9. In Figure 24-20 what is the ratio of the respective currents through the upper and the lower 15-Ω resistors?
10. Calculate the equivalent resistance between points A and B of the network in Figure 24-21. If a 50-volt battery were connected across these terminals, what current would flow?

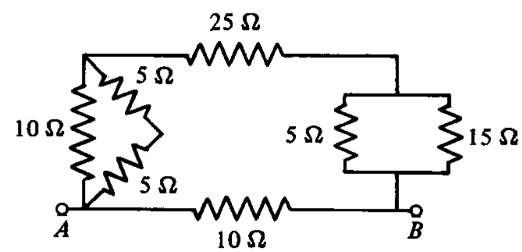


Figure 24-21

11. Twelve identical resistors, each of resistance R, are connected together as shown in Figure 24-22 to form the edges of a cube. Show that the equivalent resistance across any diagonally opposite vertices such as A and B is $5R/6$. (*Hint*: Show, by use of symmetry, that if I is the current through a battery that is connected across A and B, then the current in any of the 12 branches can only be $I/3$ or $I/6$.)

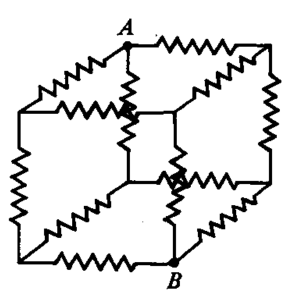


Figure 24-22

12. Suppose a thirteenth resistor also of resistance R is connected directly across the diagonal from A

to B in Figure 24-22. What is the equivalent resistance between these points now?

*13. Consider the infinite network of resistors in Figure 24-23. Show that the equivalent resistance between points A and B is

$$R = \frac{1}{2}(R_1 + R_2) + \frac{1}{2}[(R_1 + R_2)(R_1 + R_2 + 4R_3)]^{1/2}$$

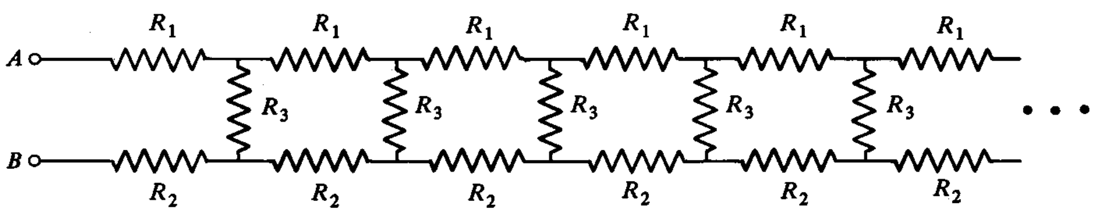


Figure 24-23

14. Consider the circuit in Figure 24-6.
   (a) How many branches and junctions are there in this circuit?
   (b) Assign a symbol and direction for the current in each branch and write down the implications of Kirchhoff's rules.
   (c) Find the value of the current in each branch, assuming that $R_1 = 20\ \Omega$, $R_2 = 10\ \Omega$, $r = 0.5\ \Omega$, and $\mathscr{E} = 15$ volts.
15. Consider the circuit in Figure 24-24.
   (a) Assign a symbol for the current in each branch and write down the implications of Kirchhoff's rules to this network.
   (b) Calculate the rate at which energy is being put out by each of the batteries.

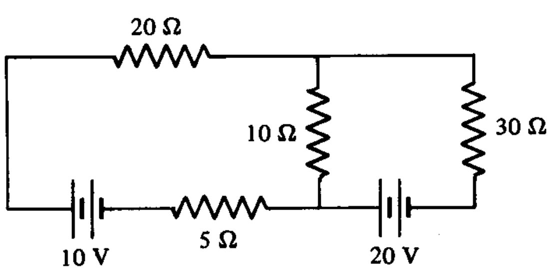


Figure 24-24

16. For the circuit in Figure 24-25 calculate the direction and magnitude of the current in the 20-Ω and 30-Ω resistors.

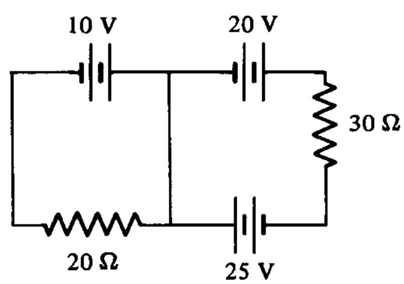


Figure 24-25

17. For the circuit depicted in Figure 24-26:
   (a) Calculate the current in the 100-Ω resistor.
   (b) Set up and solve Kirchhoff's rules to find the current through each of the batteries.
   (c) Is work being carried out *on* any of the batteries?

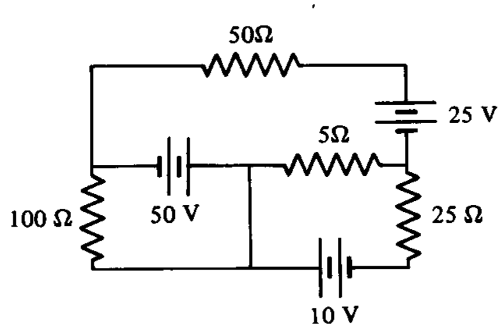


Figure 24-26

18. Without setting up Kirchhoff's rules in their entirety, for the circuit in Figure 24-27 calculate:

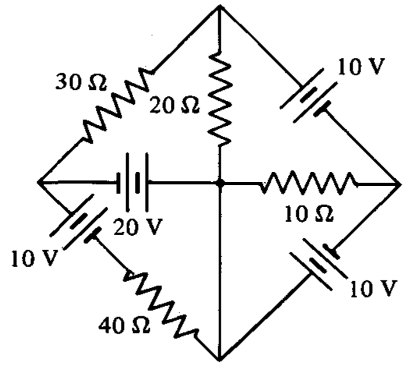


Figure 24-27

(a) The current (magnitude and direction) in the 10-Ω resistor.
(b) The current in the 40-Ω resistor.
(c) The current in the 20-Ω resistor by making use of your result to (a).
(d) The current in the 30-Ω resistor.

***19.** Consider the special Wheatstone bridge circuit depicted in Figure 24-28. Set up the implications of Kirchhoff's rules for this circuit and show that the current i through the galvanometer is

$$i = \frac{\mathscr{E}(R - R_1)}{(R_0 + 2R_G)(R + R_1) + 2RR_1}$$

where R_G is the resistance of the galvanometer.

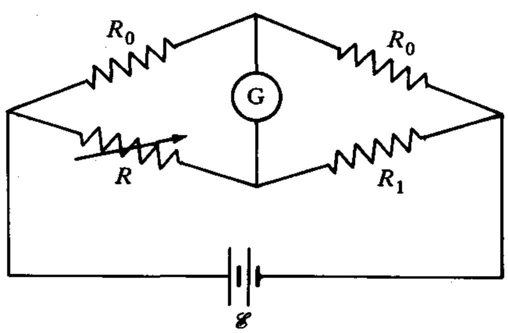


Figure 24-28

20. Consider the circuit in Figure 24-10. Carry out the analysis of this circuit, but this time assume that the resistor r and R are comparable. Assuming that $r = R = 100\ \Omega$ and $C = 10\ \mu\text{F}$, what is the time constant associated with this circuit?

21. For the circuit in Figure 24-10, assuming that $r = 2\ \Omega$, $R = 10^3\ \Omega$, $\mathscr{E} = 60$ volts, and $C = 5.0\ \mu\text{F}$, calculate:
(a) The time constant τ of the circuit.
(b) The initial current through the battery.
(c) The final charge on the capacitor.

22. Assuming that $R = 10^3\ \Omega$, $r = 1\ \Omega$, $C = 10\ \mu\text{F}$, and $\mathscr{E} = 0.5$ volt, and that just before the currents are allowed to flow the capacitor had a charge $q_0 = 2\ \mu\text{C}$, calculate the three currents, as functions of time, flowing in the circuit in Figure 24-10.

23. In Figure 24-29 if the switch S is closed at $t = 0$ and the capacitors are originally uncharged, calculate:
(a) The initial current through the battery.
(b) The final charge on each capacitor and the final value for the current through the battery.

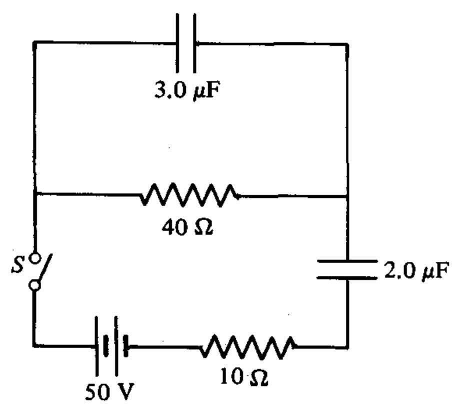


Figure 24-29

24. If the capacitors in the circuit in Figure 24-30 are originally uncharged and the switch S is closed at $t = 0$, calculate:
(a) The initial value of the current i_0 through the battery.
(b) The value of this current after a long time.

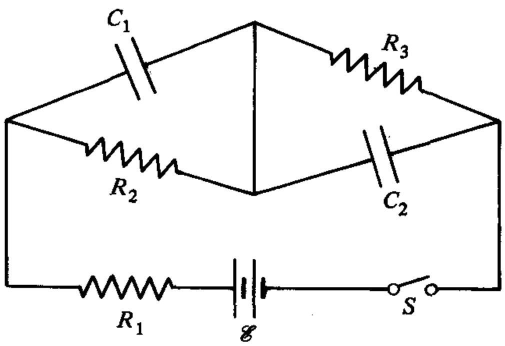


Figure 24-30

(c) The ultimate value of the charge on each capacitor.

25. Repeat both parts of Problem 23 if the 2.0-μF capacitor in Figure 24-29 has a charge of 25 μC initially. Assume that the upper plate has the positive charge initially.

26. Consider the circuit in Figure 24-12, but suppose that initially the capacitor has a charge of 50 μC. If the *left-hand* plate in the figure has the positive charge initially, calculate:

(a) The initial current in the 200-Ω resistor.

(b) The initial current through the battery.

(c) The initial current to the capacitor.

27. Repeat all parts of Problem 26, but assume that initially the right-hand plate has the positive charge.

28. For the circuit in Figure 24-31, assuming that initially the capacitors are uncharged, calculate:

(a) The initial currents through R_1, R_2, and R_3.

(b) The final charge on each capacitor.

(c) The final current through the battery.

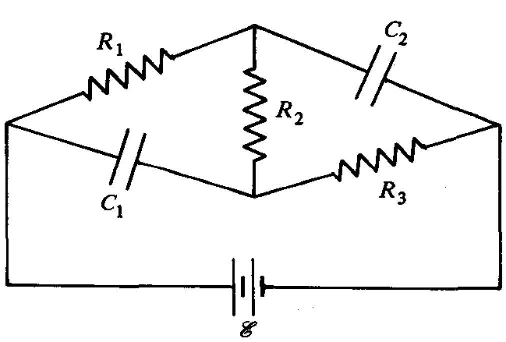


Figure 24-31

29. Repeat (a) of Problem 28, but assume this time that initially C_1 has a charge $C_1\mathscr{E}/2$ (with the left plate positive).

30. Repeat Problem 29, but assume that initially C_1 and C_2 have the respective charges Q_1 and Q_2, with the left-hand plate positive in each case.

31. Consider the circuit in Figure 24-29 and let $q_1\,(>0)$ and $q_2\,(>0)$ be the respective charges at time t on the left-hand plate of the 3.0-μF capacitor and the upper plate of the 2.0-μF capacitor.

(a) Write down Kirchhoff's rules in terms of the currents i_1, i_2, and i_3 through the battery, the 40-Ω resistor, and the 3.0-μF capacitor, respectively.

(b) What must be the initial values for i_1, i_2, and i_3, assuming that the capacitors are originally uncharged?

(c) Solve the system of equations in (a) for the currents i_1, i_2, and i_3 at any time t, and compare with your result to (b) at $t = 0$.

32. Apply Kirchhoff's rules to the circuit in Figure 24-31. Are the initial values of the capacitor charges relevant in writing down these circuit equations? At what point in the solution do the values for these charges play a role?

33. In an attempt to measure a resistance R, an ammeter and a voltmeter are connected with a battery and the resistor as shown in Figure 24-32. Show that

$$\frac{1}{R} = \frac{i}{V} - \frac{1}{R_M}$$

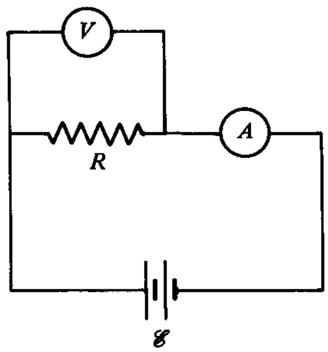


Figure 24-32

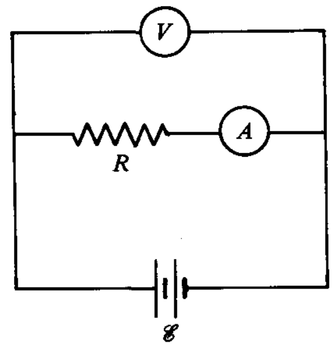


Figure 24-33

where i is the ammeter reading, V is the voltmeter reading, and R_M is the multiplier resistance. In general, $R_M \gg R$, and thus the ratio i/V is a good measure of R.

34. Consider the circuit in Figure 24-33. Show that if the shunt resistance R_s of the ammeter is negligible compared to the galvanometer resistance, then

$$R = \frac{V}{i} - R_s$$

where V/i is the ratio of the voltmeter reading to the ammeter reading.

25 The magnetic field

In amber there is a flammeous and spiritous nature and this by rubbing on the surface is emitted by hidden passages and does the same that lodestone does.

PLUTARCH

25-1 Introduction

The very existence of the word *electromagnetism* implies that there must be a connection of some type between electric and magnetic phenomena. In order to explore the nature of this relationship, in this and the next two chapters we turn from our studies of purely electrical effects to a consideration of physical systems that exhibit magnetic behavior.

As noted in Chapter 19, the phenomenon of magnetism was discovered in antiquity. The word "magnetism" itself is derived from the name for a certain region in Asia Minor, *Magnesia,* where the iron ore $FeO\text{-}Fe_2O_3$, or lodestone (leading stone), occurred in abundance. These stones were found to be able to exert attractive and repulsive forces on each other—depending on their relative orientation. Further, they could also impart their magnetic property to other nearby iron and steel objects by *magnetizing* them. Today we know that the earth itself is a huge magnet and is capable of magnetizing iron and steel rods. Also well known is the related fact that a thin, magnetized needle, which is free to rotate about a vertical axis through its

center, will invariably orient itself along the north-south direction. This distinctive behavior of a magnetized needle was discovered in the twelfth century and is basic to the operation of the mariner's compass.

25-2 Magnetic poles

Since early in the nineteenth century it has been known that magnetic effects can be produced by electric currents. Our discussion of magnetism will reflect this fact throughout. By way of an introduction to these ideas, in this section we review briefly some of the earlier known and thus more qualitative aspects of magnetism.

Consider a thin, magnetized iron needle, which is suspended so that it is free to rotate in a horizontal plane. As in the mariner's compass and as shown in Figure 25-1, the needle experiences a torque about its center that causes it to rotate until it is lined up along the north-south direction. That end of the needle pointing north is said to be its *north* (or north-seeking) *pole* and the other end is known as its *south* (or south-seeking) *pole*. Once the north pole of a magnetized needle has been identified and suitably marked, if suspended, the needle will invariably point in a northerly direction from any point on the earth. Two notable exceptions to this are the regions near the two magnetic poles of the earth itself.

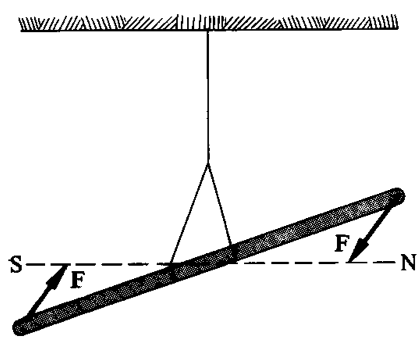


Figure 25-1

Figure 25-2a shows two thin, magnetized needles, suspended so that they are free to rotate in a plane (the plane of the page in the figure) about their respective centers. Experiment shows that, just as for the case of electric charge, a north and south pole attract each other and that two north poles or two south poles repel each other. Further, just as for electric charge, the strength of these forces also decreases as the separation distance between the poles increases and, as shown in the figure, the force between any two poles lies along the line joining them. As a consequence of these forces, each of the needles experiences a torque, which tends to line it up as in Figure 25-2b. Only in this latter configuration is there no torque acting on either needle.

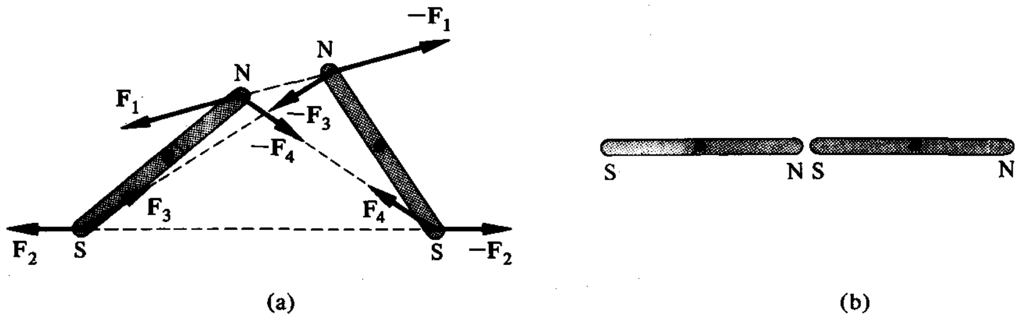


Figure 25-2

By analogy to our studies of electrostatics, the above properties of magnetic poles suggest that we study magnetism by analyzing these forces between magnetic poles in quantitative terms. The main reason this is not usually done has to do with the fact that it is *impossible to isolate a magnetic pole completely.* This is often described by saying that *magnetic monopoles* do not exist. To obtain an experimental verification of this feature consider, in Figure 25-3a, a thin, magnetized needle with its two poles. On cutting this needle in two we find, as in Figure 25-3b, that instead of physically separating the original two poles, a new north and south pole appear at the cut surfaces. Thus two magnetized needles, each with its own north and south pole, result. Even if this subdivision process is continued down to a microscopic level, an isolated pole never appears.

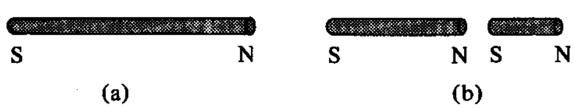


Figure 25-3

It is worth emphasizing that magnetic forces are very much distinct from electrostatic forces. If, for example, a charged body is brought near a magnetized rod, no force arises between them except possibly for that due to any charge that may be induced on the rod. However, as will be seen in the next section, there is an unambiguous force between a charged body and a magnetized rod *if they are in relative motion.*

25-3 The discoveries of Oersted and Ampere

After the invention of the voltaic cell by Alessandro Volta (1745–1827), it became possible for the first time to produce steady electric currents at will, and to study phenomena associated with them. In 1820 Hans Christian Oersted (1777–1851) discovered that a wire in which there flows such a current has properties similar to that of a permanent magnet. That is, he found that a wire carrying a current behaves as if it had been magnetized and that this magnetic property ceases when the current does. Not unexpectedly,

the study of magnetism received a tremendous impetus by this work. Thereafter, one important discovery followed the other in rapid succession. Today it is generally recognized that all observed magnetic effects are due to one of two basic sources. These are: (1) the motion of electric charge as in an electric current; and (2) certain intrinsic magnetic properties of the microscopic constituents of matter, particularly those associated with a property of the electron known as its *spin*. In the following, we shall be concerned mainly with describing magnetic effects produced by the motion of electric charge. To understand the magnetic effects associated with the spin of the electron, some knowledge of quantum mechanics is required; therefore further consideration of this matter will be postponed at this point.

Consider, in Figure 25-4, a tightly wound coil of wire, or solenoid, carrying a current i. Experiment shows that this coil behaves in many respects like the magnetized iron rods or needles considered in the previous section, and that its magnetic behavior becomes more pronounced the greater is the current. For the assumed direction of the current, as shown in the figure, the ends of the coil take on the attributes of a north and a south pole. If, for example, a magnetized needle is brought near the coil, it will rotate in the direction shown. The situation here is very much analogous to that in Figure 25-2.

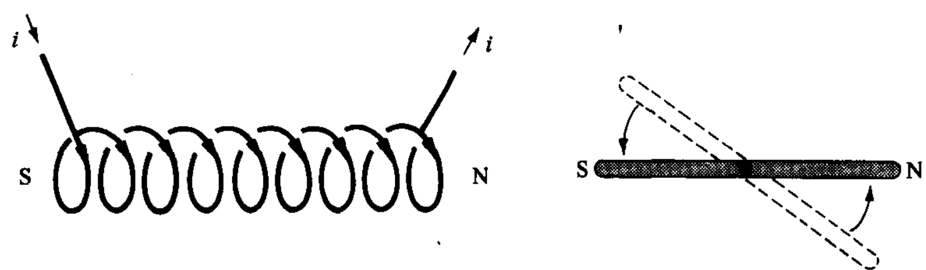


Figure 25-4

Since currents appear to have magnetic properties, we might expect that two currents will exert forces on each other. The fact that this is indeed the case was first established, shortly after Oersted's discovery, by André Marie Ampère (1775–1836). Figure 25-5 shows schematically how such an experiment may be carried out. Two stiff wires are connected across a seat of electromotive force and allowed to hang freely in a pool of liquid mercury. They are thus in electrical contact and can at the same time move relative to each other. Figure 25-5a shows the situation when the switch S is open, so that there is no current. The wires hang in a vertical position along the force of gravity in this case. If the switch S is closed, then, as in Figure 25-5b, certain currents i will flow in opposite directions through the two wires. As shown, the two wires exert repulsive forces on each other in this circumstance. Similar experiments can be used to establish that parallel currents attract each other.

It is also interesting to confirm that the force between currents is due to

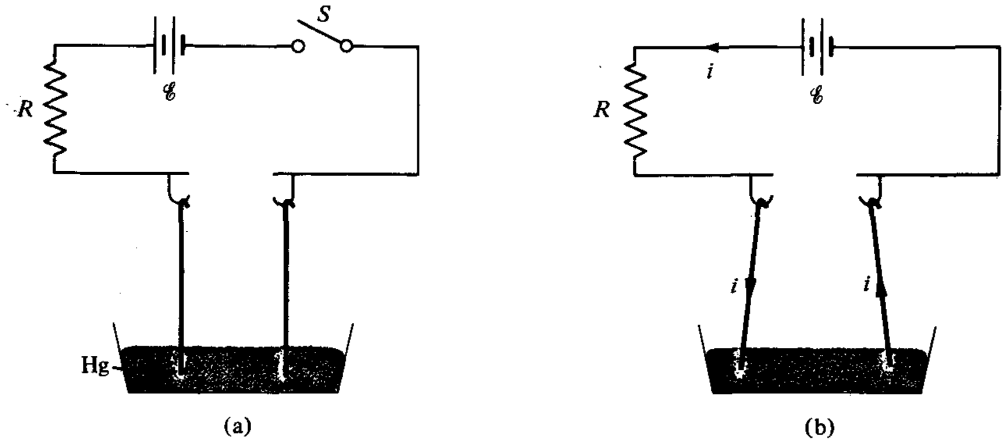


Figure 25-5

the fact that charges are in relative motion and not to any of the material properties of the wires. For this purpose, it is convenient to carry out an experiment by use of an *oscilloscope,* or *cathode-ray tube.* The main elements of such a tube are an electron gun, and directional controlling electric plates by means of which an electron stream can be sent down the tube to make a spot on the screen at its lower end. Figure 25-6a shows schematically such a tube, in which an electron beam makes a spot on the center of the screen. If now, as in Figure 25-6b, an upward current i opposite to the original direction of the electron stream is generated, the spot is deflected to the right. Similarly, if as in Figure 25-6c the external current i flows downward, the electron stream will be deviated to the left. Recalling that electrons have a negative charge, so that a downward flow of electrons corresponds to an upward flow of current, these deviations of the electron beam are easily accounted for. For example, in Figure 25-6c the two currents flow in opposite directions and thus repel each other, just as in Figure 25-5b.

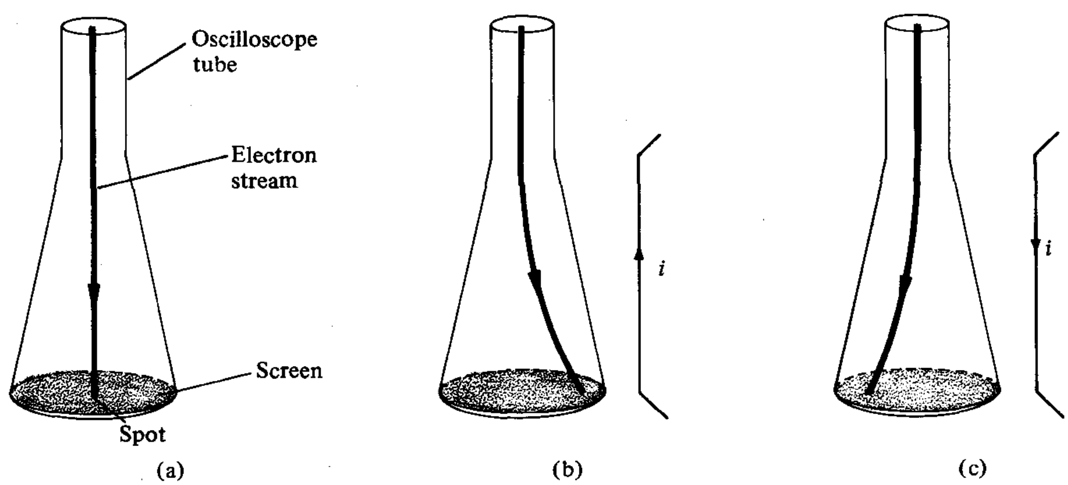


Figure 25-6

25-4 The magnetic field

In our studies of electrostatics it was convenient to think of the region of space near charged particles to be modified by virtue of the existence of an electric field **E** in that region. This field **E** is a vector field and represents at each point the force per unit charge on a particle placed at that point.

Similiarly, it is convenient to think of a *magnetic field* as existing at each point of space near a collection of magnetic sources such as currents or magnetized bodies. We say that a magnetic field exists at such a point if a force, other than that due to an electric field, is produced on a charged particle *moving* through this point. As for the electric field, the magnetic field is a vector field and associates with each point in space both a magnitude and a direction. For the present, this field will be specified by a certain vector function **B**, which is known as the *magnetic induction field*[1] or the **B**-*field*. The purpose of this section is to define this field and to discuss some of its properties.

Imagine a region of space containing various magnetic sources. Experiments involving the observation of the trajectories of charged particles traveling through this region show that the force **F** acting on these particles has the following properties:

1. **F** is directly proportional to the charge q of the particle.
2. **F** is directly proportional to the magnitude v of the velocity **v** of the particle.
3. **F** is perpendicular to **v** throughout the trajectory of the particle.

Thus, doubling the charge or the velocity of a particle results in the doubling of the strength of the magnetic force. The third property of **F**, that it is invariably at right angles to the particle's velocity, is perhaps its most unusual property. It rests on the fact that the kinetic energy of a particle moving in a **B**-field does not vary throughout its motion. Hence **F** must always be at right angles to **v**. For, if **F** had a component along **v**, then work would be carried out on the particle and its kinetic energy would change according to the work-energy theorem. The fact that the kinetic energy of the particle does not vary thus confirms the fact that **F** is invariably perpendicular to **v**.

Because of the above experimental properties of the magnetic force **F**, we can define the magnetic induction field **B** associated with the given sources by the relation

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \tag{25-1}$$

where the symbol $\mathbf{v} \times \mathbf{B}$ stands for the *cross product* of the vectors **v** and **B**. According to the definition in Section 10-3, the cross product $\mathbf{v} \times \mathbf{B}$ is

[1]For historical reasons, the term *magnetic field* is reserved for a different quantity, which is, however, directly related to the **B**-field. It will be defined in Chapter 28.

perpendicular to both **v** and **B** and thus, as required by experiment, **F** is perpendicular to **v**. Further, since experiment shows **F** to be directly proportional to both **q** and $|\mathbf{v}|$, it follows from (25-1) that the **B**-field as defined here is independent of both **q** and **v**; *it is determined exclusively by the magnetic sources.*

Figure 25-7 summarizes the relations between the directions of the three vectors **v**, **B**, and **F**. Assuming that **B** and **v** lie in the x-y plane, **F** will be parallel to the z-axis. As shown, **F** points along the positive sense of the z-axis for $q > 0$ and in the opposite direction for $q < 0$. The magnitude F of this force is $|qvB \sin \theta|$ with θ ($< 180°$) the angle between **v** and **B**.

It follows from (25-1) that the unit of **B** is the newton second per coulomb meter (N-s/C-m), or the newton per ampere meter (N/A-m). Let us define the unit of the tesla (abbreviated T) by

$$1\ \mathrm{T} = 1\ \mathrm{N/A\text{-}m}$$

so that the unit of the **B**-field is the tesla. A related unit is the weber (Wb). It is defined by

$$1\ \mathrm{Wb} = 1\ \mathrm{T\text{-}m^2} \tag{25-2}$$

and thus 1 $\mathrm{Wb/m^2}$ is an equivalent unit for **B**. Parenthetically, let us note the unit of the gauss (G), which is defined by

$$1\ \mathrm{G} = 10^{-4}\ \mathrm{T} = 10^{-4}\ \mathrm{Wb/m^2}$$

but which will not be used in this book. The **B**-field of the earth, for example, has at its surface a magnitude of the order of 5×10^{-5} tesla or, equivalently, 0·5 gauss.

Just as for the electric field, it is convenient to imagine **B**-field lines to exist in the space around magnetic sources. Each of these field lines is defined so that the tangent to a line at a given point is parallel to **B** at that point. Also, the density of field lines in any region is proportional to the magnitude of **B** in that region. By way of illustration, Figure 25-8 shows the field lines about

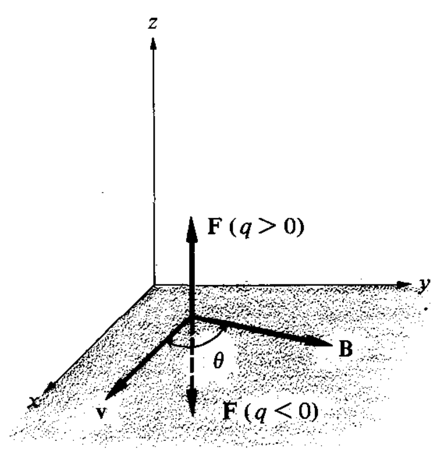


Figure 25-7

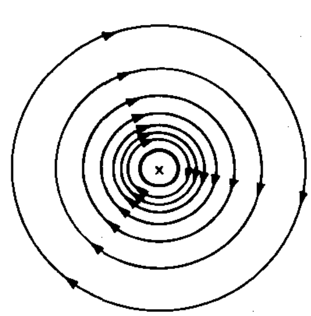


Figure 25-8

a long, straight wire with current directed perpendicularly down into the plane of the diagram. Note that the field lines are circles, concentric with the axis of the wire, and that their density decreases with increasing distance from the current. Figure 25-9 shows some of the field lines produced by the current in a circular loop of wire.

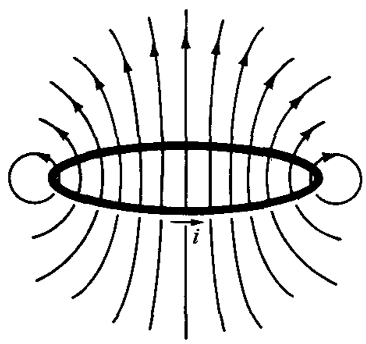


Figure 25-9

Example 25-1 Consider a proton traveling due east in the upper reaches of the atmosphere at a velocity of 5.0×10^5 m/s. Assuming the earth's field at this point has a strength of 0.5×10^{-4} tesla and is directed due south, calculate:
(a) The magnitude of the force on the proton.
(b) The direction of this force.
(c) The acceleration of the proton.

Solution
(a) Since **v** and **B** are at right angles, according to (25-1) the magnitude of **F** is

$$F = qvB = (1.6 \times 10^{-19}\,\text{C}) \times (5.0 \times 10^5\,\text{m/s}) \times (0.5 \times 10^{-4}\,\text{T})$$
$$= 4.0 \times 10^{-18}\,\text{N}$$

(b) From the facts that the charge on the proton is positive, and **v** is directed east while **B** is south, it follows by reference to Figure 25-7 or to the definition of a cross product and (25-1) that the force on the proton is directed vertically downward.
(c) According to Newton's law, the acceleration of the proton is the ratio F/m. Using the known value $m = 1.67 \times 10^{-27}$ kg, and the above value for F, there results

$$a = \frac{F}{m} = \frac{4.0 \times 10^{-18}\,\text{N}}{1.67 \times 10^{-27}\,\text{kg}} = 2.4 \times 10^9\,\text{m/s}^2$$

Since the proton has a positive charge, the direction of this acceleration is also directed downward.

Example 25-2 Show explicitly that if a particle of charge q and mass m moves in a static magnetic induction **B**, then its kinetic energy is constant in time.

Solution Making use of the given data, (25-1), and Newton's law, we have

$$m\frac{d\mathbf{v}}{dt} = q\mathbf{v} \times \mathbf{B}$$

where $d\mathbf{v}/dt$ is the acceleration of the particle at the instant when its velocity is **v**.

Taking the dot product of this relation with **v**, we find that

$$m\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}=q\mathbf{v}\cdot(\mathbf{v}\times\mathbf{B})$$

But by definition of a cross product, the vector $\mathbf{v}\times\mathbf{B}$ is at right angles to **v**, and thus according to the definition of the dot product the right-hand side vanishes. Hence

$$m\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}=0$$

and since this can be expressed in the equivalent form

$$\frac{d}{dt}\left(\frac{m}{2}v^2\right)=0$$

the desired result follows.

25-5 The Biot-Savart law

In this section we consider the problem of calculating the **B**-field associated with current flows in *thin* wires by use of the experimentally based law of *Biot-Savart.* The problem of calculating **B** for other current distributions will be discussed in Sections 25-8 and 25-9.

Consider, in Figure 25-10, an infinitesimal current element of length $d\mathbf{l}$ directed parallel to the current flow i in a thin wire. The law of Biot-Savart states that the magnetic induction $d\mathbf{B}$ at a point P at the vectorial displacement **r** from this current element is

$$d\mathbf{B}=\frac{\mu_0}{4\pi}i\frac{d\mathbf{l}\times\mathbf{r}}{r^3} \tag{25-3}$$

where μ_0 is a constant defined by

$$\mu_0=4\pi\times10^{-7}\ \text{Wb/A-m}$$

Recalling the definition of a cross product, we see from (25-3) that the magnitude of $d\mathbf{B}$ is

$$|d\mathbf{B}|=\frac{\mu_0}{4\pi}i\frac{|d\mathbf{l}|\sin\theta}{r^2} \tag{25-4}$$

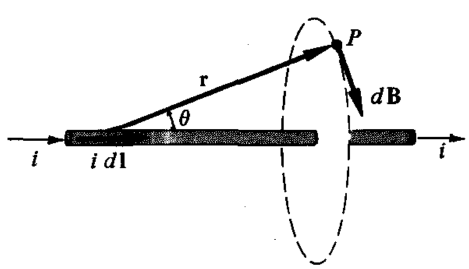


Figure 25-10

with θ the angle between $d\mathbf{l}$ and $\mathbf{r}$. The direction of $d\mathbf{B}$ may be obtained, as in the figure, by drawing a circle through the point P so that $d\mathbf{l}$ lies on its axis. The vector $d\mathbf{B}$ is invariably tangent to this circle. The *sense* of $d\mathbf{B}$ may be obtained by use of the right-hand rule: If the current element $d\mathbf{l}$ is grasped in the right hand, with the outstretched thumb pointing along the direction of the current, then the fingers will be oriented along the direction of $d\mathbf{B}$.

The implications of (25-3), as far as the direction of $d\mathbf{B}$ is concerned, are illustrated in Figure 25-11. Since $d\mathbf{B}$ is always at right angles to both $d\mathbf{l}$ and $\mathbf{r}$, the results in the figure follow directly either by use of the definition of the cross product or from the right-hand rule. The fact that $d\mathbf{B}$ vanishes at all points along the y-axis is a consequence of (25-3), since for these points $\mathbf{r}$ and $d\mathbf{l}$ are parallel vectors, and for these the cross product always vanishes.

Just as for the electrostatic field, experiment shows that the *superposition principle* is also valid for the **B**-field. That is, the magnetic induction due to two or more current elements is the vector sum of the magnetic inductions produced by the individual elements separately. Hence the magnetic induction **B** produced by the current in a closed circuit is

$$\mathbf{B}=\frac{\mu_0}{4\pi}\oint\frac{i\,d\mathbf{l}\times\mathbf{r}}{r^3} \tag{25-5}$$

where the integral is to be carried out over every element $d\mathbf{l}$ of the closed loop in which the current i flows. Since no truly isolated current element $i\,d\mathbf{l}$ exists, (25-5) is really more useful, and it is for this latter form that the term the *Biot-Savart law* is generally reserved.

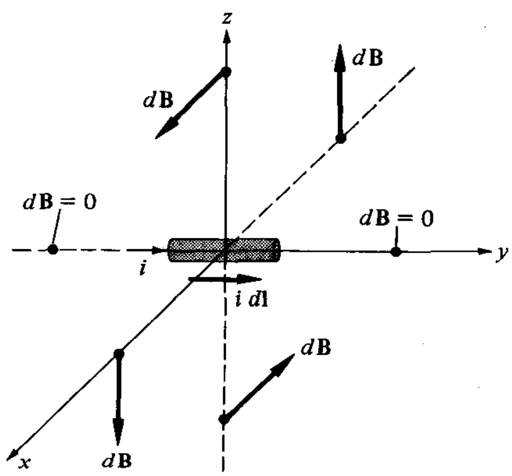


Figure 25-11

25-6 Applications of the Biot-Savart law

The purpose of this section is to derive the **B**-field associated with a number of simple current distributions. Specifically, we shall establish the following:

1. The **B**-field at a perpendicular distance r from a very long, straight wire carrying a current i has the magnitude

$$B = \frac{\mu_0 i}{2\pi r} \tag{25-6}$$

and its direction is tangent to the circle of radius r with center on the wire. The sense of **B** is shown in Figure 25-12 and is given by the same "right-hand" rule as above for a small current element $i\,d\mathbf{l}$:

Grasp the wire in the right hand with the thumb pointing along the direction of the current. The fingers will then circle the wire in the sense of the direction of the **B***-field.*

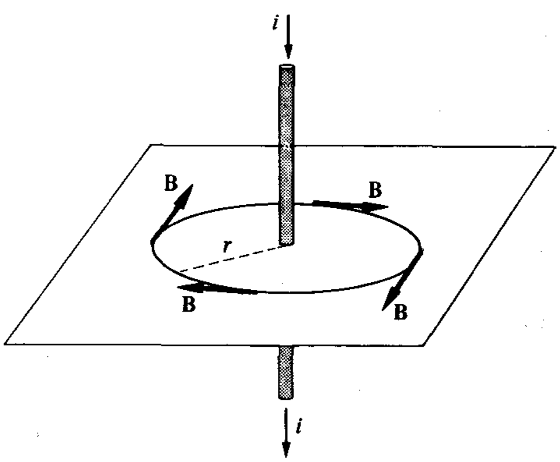


Figure 25-12

2. The magnitude of the **B**-field at a field point on the axis of a circular loop of wire of radius a and carrying a current i is

$$B = \frac{\mu_0 i a^2}{2(a^2 + b^2)^{3/2}} \tag{25-7}$$

where b is the distance from the center of the loop to the field point. See Figure 25-13. The direction of **B** is parallel to the axis of the loop and its sense is given by a rule also known as a right-hand rule. It may be obtained from the preceding one associated with the current in a long wire by interchanging the roles of i and **B**:

Grasp the loop in the right hand with the fingers pointing along the direction of the current. The thumb then points along **B***.*

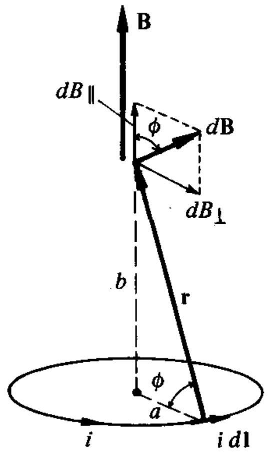


Figure 25-13

3. The magnitude of the **B**-field at a point P *on the axis* of a tightly wound solenoid carrying a current i and containing n turns per unit length is

$$B = \frac{\mu_0 n i}{2}[\cos \alpha_1 + \cos \alpha_2] \tag{25-8}$$

The angles α_1 and α_2, as defined in Figure 25-14, are the angles between the axis of the solenoid and the rays from P to the edges of the solenoid. The circles with crosses represent current going perpendicularly down into the plane of the diagram and correspondingly circles with dots represent current coming out of the diagram. The direction of **B** is parallel to the axis of the solenoid, and its sense is determined by the above right-hand rule of a single loop. For the special case of a very long solenoid, $\alpha_1 = \alpha_2 \approx 0$, the result in (25-8) reduces to

$$B = \mu_0 n i \tag{25-9}$$

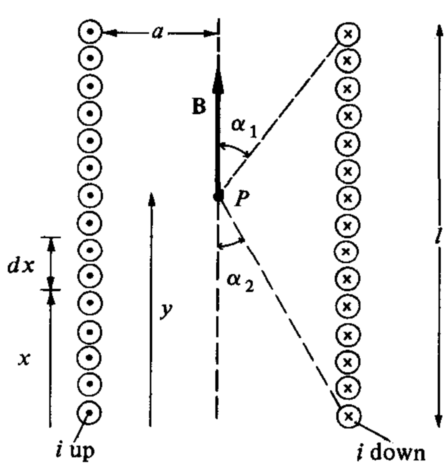


Figure 25-14

The validity of (25-6) through (25-8) is established below in Examples 25-3 through 25-5, respectively.

Example 25-3 Establish (25-6) by use of the Biot-Savart law.

Solution As shown in Figure 25-15, let us set up a coordinate system with the current along the z-axis and with the field point P on the x-axis a distance x_0 from the origin. The magnetic induction $d\mathbf{B}$ due to a current element $i\,dz$ at a distance z from the origin is perpendicular to the x-z plane and points along the positive y-axis according to the Biot-Savart law. (See also Figure 25-10 in this connection.) According to (25-4), its magnitude, dB, is

$$dB = \frac{\mu_0 i}{4\pi}\frac{dz}{r^2}\sin\theta = \frac{\mu_0 i}{4\pi}\frac{dz}{(z^2+x_0^2)}\frac{x_0}{(z^2+x_0^2)^{1/2}}$$
$$= \frac{\mu_0 i}{4\pi} x_0 \frac{dz}{[z^2+x_0^2]^{3/2}}$$

where the second equality follows since $r = (x_0^2+z^2)^{1/2}$ and $\sin\theta = \sin(\pi-\theta) = x_0/(x_0^2+z^2)^{1/2}$. The total field is obtained by integrating over all values of z from $-\infty$ to $+\infty$. The result is

$$B = \int dB = \frac{\mu_0 i x_0}{4\pi}\int_{-\infty}^{\infty}\frac{dz}{[z^2+x_0^2]^{3/2}} = \frac{\mu_0 i x_0}{4\pi}\left[\frac{1}{x_0^2}\frac{z}{[z^2+x_0^2]^{1/2}}\right]\Bigg|_{-\infty}^{\infty}$$
$$= \frac{\mu_0 i}{4\pi x_0}[1-(-1)] = \frac{\mu_0 i}{2\pi x_0}$$

where the third equality follows by consulting a table of integrals. Finally, identifying the distance x_0 with the radial distance r, the validity of (25-6) is established.

Example 25-4 Establish the validity of (25-7) by use of the Biot-Savart law.

Solution Consider, in Figure 25-13, the field $d\mathbf{B}$ due to a current element $i\,d\mathbf{l}$ of the loop. As shown, $d\mathbf{B}$ is perpendicular to both $d\mathbf{l}$ and $\mathbf{r}$ in accordance with (25-3). Let its components along and perpendicular to the axis be $dB_\parallel$ and $dB_\perp$, respectively. By symmetry, the sum of the contributions to $dB_\perp$, as we integrate around the loop, will

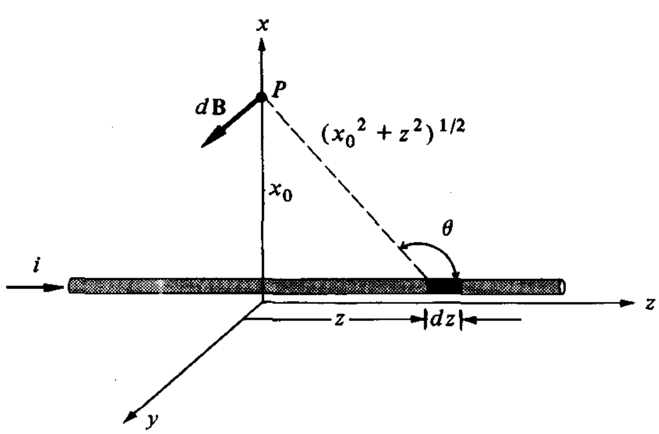


Figure 25-15

cancel. Hence the total field **B** will be directed along the axis of the loop and can be obtained by integrating $dB_\parallel$ around the loop.

To determine B let us first note that

$$dB_\parallel = |d\mathbf{B}| \cos\phi = |d\mathbf{B}| \frac{a}{(a^2+b^2)^{1/2}}$$

where ϕ is as defined in the figure. Since the angle θ between $d\mathbf{l}$ and $\mathbf{r}$ is here $\pi/2$, and since $r = (a^2+b^2)^{1/2}$, it follows by use of (25-4) that

$$dB_\parallel = \frac{\mu_0 i\, dl}{4\pi} \frac{1}{(a^2+b^2)} \frac{a}{(a^2+b^2)^{1/2}} = \frac{\mu_0 i a\, dl}{4\pi[a^2+b^2]^{3/2}}$$

The total field B is obtained by integrating $dB_\parallel$ over the entire loop:

$$\begin{aligned} B = \int dB_\parallel = \int \frac{\mu_0 i a}{4\pi[a^2+b^2]^{3/2}}\, dl &= \frac{\mu_0 i a}{4\pi[a^2+b^2]^{3/2}} \int dl \\ &= \frac{\mu_0 i a(2\pi a)}{4\pi[a^2+b^2]^{3/2}} \\ &= \frac{\mu_0 i a^2}{2[a^2+b^2]^{3/2}} \end{aligned}$$

where the fourth equality follows since the length of the loop is $2\pi a$. The validity of (25-7) is thereby established.

Example 25-5 Derive (25-8) for the **B**-field on the axis of a solenoid. As shown in Figure 25-14, assume that it consists of N turns, has a radius a, length l, and $n \equiv N/l$ turns per unit length.

Solution To calculate the field at a point P on the axis and at a distance y from one end of the solenoid, let dB be the field at P due to $(n\,dx)$ turns at a distance x from the same end. See Figure 25-14. According to (25-7), dB is

$$dB = \frac{\mu_0 i a^2}{2[a^2+(y-x)^2]^{3/2}}\, n\, dx$$

since in the present case $b = y - x$. Integrating over all values of x from 0 to l yields for the total field

$$\begin{aligned} B = \int dB &= \frac{\mu_0 i a^2 n}{2} \int_0^l \frac{dx}{[a^2+(y-x)^2]^{3/2}} \\ &= \frac{\mu_0 i a^2 n}{2} \frac{1}{a^2} \left[\frac{x-y}{[a^2+(y-x)^2]^{1/2}}\right]\Bigg|_0^l \\ &= \frac{\mu_0 i n}{2} \left[\frac{l-y}{[a^2+(l-y)^2]^{1/2}} + \frac{y}{[a^2+y^2]^{1/2}}\right] \end{aligned}$$

Reference to Figure 25-14 shows that $\cos\alpha_1 = (l-y)/[a^2+(l-y)^2]^{1/2}$ and $\cos\alpha_2 = y/[y^2+a^2]^{1/2}$. The validity of (25-8) is therefore established.

The direction of **B** may be determined by the right-hand rule. For the assumed sense of the current in Figure 25-14 (recall that crosses represent current flow perpendicularly down and dots represent current flow up), **B** is directed as shown.

25-7 Further applications

To obtain a feeling for orders of magnitude in this section we shall apply some of the above results to particular cases.

Example 25-6 Two very long, parallel wires are 0.5 meter apart and carry currents of 3.0 amperes in opposite directions, as shown in Figure 25-16. Calculate the **B**-field at a point between the wires and at a perpendicular distance of 0.4 meter from one of them.

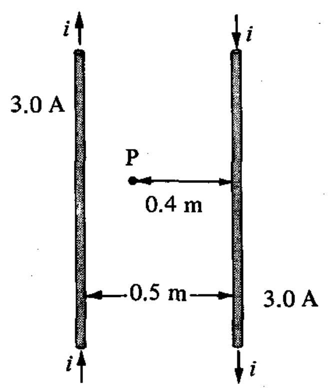


Figure 25-16

Solution The field $\mathbf{B}_L$ at P due to the wire on the left is directed perpendicularly down into the plane of Figure 25-16. Its magnitude is

$$B_L = \frac{\mu_0 i}{2\pi r} = \frac{(4\pi \times 10^{-7}\ \mathrm{Wb/A\text{-}m}) \times (3.0\ \mathrm{A})}{2\pi \times 0.1\ \mathrm{m}}$$
$$= 6.0 \times 10^{-6}\ \mathrm{T}$$

Correspondingly, the field $\mathbf{B}_R$ at P due to the other current is also directed perpendicularly down and it has the magnitude

$$B_R = \frac{\mu_0 i}{2\pi r} = \frac{(4\pi \times 10^{-7}\ \mathrm{Wb/A\text{-}m}) \times (3.0\ \mathrm{A})}{2\pi \times 0.4\ \mathrm{m}}$$
$$= 1.5 \times 10^{-6}\ \mathrm{T}$$

The total field $(\mathbf{B}_L + \mathbf{B}_R)$ is therefore also directed perpendicularly down and its magnitude B is

$$B = B_L + B_R = 6.0 \times 10^{-6}\ \mathrm{T} + 1.5 \times 10^{-6}\ \mathrm{T}$$
$$= 7.5 \times 10^{-6}\ \mathrm{T}$$

Example 25-7 Calculate the **B**-field at the center of a loop of radius 10 cm. Assume a current of 5.0 amperes.

Solution The direction of the field is perpendicular to the plane of the loop and its sense is given by the right-hand rule. Its magnitude is obtained by substituting into

(25-7) the parameter values $i = 5.0$ amperes, $a = 0.1$ meters, and $b = 0$. The result is:

$$B = \frac{\mu_0 i a^2}{2(a^2 + b^2)^{3/2}} = \frac{(4\pi \times 10^{-7}\ \text{Wb/A-m}) \times (5.0\ \text{A})}{2 \times 0.1\ \text{m}}$$
$$= 3.1 \times 10^{-5}\ \text{T}$$

Example 25-8 Two circular loops, each of radius a, carry parallel currents each of strength i. Assuming they are at a separation distance $2c$, calculate the magnetic induction on the axis of the loops and at a point a distance z from one of the loops. See Figure 25-17.

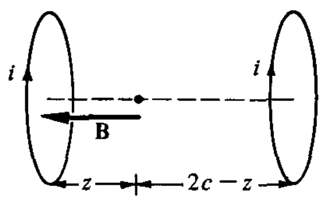


Figure 25-17

Solution Applying the right-hand rule, we see that the **B**-field is directed toward the left in the figure. Its magnitude is obtained by applying (25-7) twice, with the following result:

$$B = \frac{\mu_0 i a^2}{2}\left\{\frac{1}{(a^2 + z^2)^{3/2}} + \frac{1}{[a^2 + (2c - z)^2]^{3/2}}\right\}$$

It is established in the problems that for the choice $z = c = a/2$, the field at this point on the axis and midway between the two loops is relatively uniform. When used with these parameter values this system defines an important laboratory tool known as a *Helmholtz coil.*

Example 25-9 A solenoid has a radius of 2 cm, a length of 10 cm, and 500 turns. Assuming a current of 5.0 amperes, calculate the magnitude of **B** along the axis at:
(a) The center of the coil.
(b) One end of the coil.

Solution
(a) Since there are 500 turns in 10 cm, it follows that the number of turns per unit length is

$$n = \frac{N}{l} = \frac{500}{0.1\ \text{m}} = 5 \times 10^3/\text{m}$$

Reference to Figure 25-14 shows that at the center

$$\cos \alpha_1 = \cos \alpha_2 = \frac{5}{\sqrt{5^2 + 2^2}} = 0.93$$

Substitution into (25-8) yields

$$B = \frac{\mu_0 n i}{2}[\cos \alpha_1 + \cos \alpha_2]$$
$$= \frac{(4\pi \times 10^{-7}\ \text{Wb/A-m}) \times (5 \times 10^3/\text{m}) \times 5\ \text{A}}{2} \times [0.93 + 0.93]$$
$$= 2.9 \times 10^{-2}\ \text{T}$$

(b) At an end of the solenoid,

$$\cos\alpha_1 = 0 \qquad \cos\alpha_2 = \frac{10}{\sqrt{10^2+2^2}} = 0.98$$

since $\alpha_1 = 90°$ in this case. Proceeding as above with these values, we find that

$$B = 1.5\times 10^{-2}\,\mathrm{T}$$

25-8 Gauss' law for magnetism

As we saw in our studies of electrostatics, Gauss' law and the existence of a potential function determine, to a large extent, all of the essential features of the electrostatic field. There are two analogous laws, called *Gauss' law for magnetism* and *Ampère's law,* which play the same role for the **B**-field. The purpose of this and the next section is to discuss these two very important characterizations of the magnetic induction field.

As will be seen in the following chapters, the importance of these two laws is due, in the main, to the fact that they are basic to Maxwell's equations. Indeed, Gauss' law for magnetism is one of these four basic relations. Further, for any given distribution of currents, the laws of Ampère and of Gauss when taken together comprise a complete specification of the **B**-field everywhere. Hence they constitute the necessary generalization of the more restricted Biot-Savart law, which applies only to current flows in thin wires.

By analogy to the definition of electric flux in (20-13), we define the magnetic flux Φ_m through a surface S by

$$\Phi_m = \int_S \mathbf{B}\cdot d\mathbf{S} \tag{25-10}$$

where $d\mathbf{S}$ is an area element normal to S, $\mathbf{B}$ is the value of the **B**-field at that point and the integral is over the surface S. In terms of this quantity, Gauss' law for magnetism states that the *magnetic flux out of every closed surface vanishes.* Thus

$$\oint_S \mathbf{B}\cdot d\mathbf{S} = 0 \tag{25-11}$$

with $d\mathbf{S}$ a vectorial area element directed *outward* from the *closed* surface S. Comparison with Gauss' law for the **E**-field in (20-15) leads to the conclusion that there is no magnetic analogue of electric charge. We often describe this by saying that there are no *magnetic monopoles.* The validity of (25-11) has been established by a vast number of experiments and despite continuing research no one has as yet ever detected the presence of a magnetic monopole.

One of the important consequences of (25-11) is that all **B**-field lines must be continuous. The reasoning here is precisely the same as that referred to in Chapter 20 to establish the continuity property of the **E**-field lines in a charge-free region. Since there seem to be no magnetic monopoles, the **B**-field

lines must be continuous everywhere. Hence, the property of **B**-field lines, as exemplified in Figure 25-8, of always closing on themselves is true in general.

25-9 Ampère's law

Consider, in Figure 25-18, an open surface S bounded by a closed curve l. Let us assign a positive sense of direction along this curve by constructing an infinitesimal tangent vector $d\mathbf{l}$ at an arbitrary point of l. As viewed from the top, then, the positive sense of the curve in Figure 25-18 is counterclockwise. Associated with this sense for l let us also assign an *outward sense* to every area element $d\mathbf{S}$ of the surface S bounded by l, in accordance with the following rule:

> *Grasp the curve l in the right hand with the fingers pointing along the predefined sense of l; the outstretched thumb will then point in the direction of the outward sense of the area elements $d\mathbf{S}$ of S.*

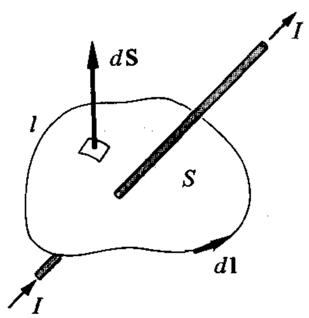


Figure 25-18

According to this rule, for example, the infinitesimal vectorial area element $d\mathbf{S}$ in Figure 25-18 points along the outward sense of S. Note that the surface S in this context is always *open*, by virtue of the fact that it is bounded by a closed curve. The assignment of a sense to this bounding curve will always be assumed in the following to associate the above outward sense for this surface.

Consider now an orbitrary open surface S with its bounding curve l in a region of space through which flow certain electric currents. *Ampère's* law states that the components of the **B**-field along the curve l are related to the net current I, which flows through S along its outward sense, by

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0 I \tag{25-12}$$

where the integral is a *line integral* (Section 21-2) and represents the sum of the products $\{\mathbf{B} \cdot d\mathbf{l}\}$ for all elements $d\mathbf{l}$ of the closed curve l. It is important to note that (25-12) applies to *all* closed curves l, and that for a given l it is

applicable to all open surfaces S bounded by that curve. Also implicit in (25-12) is the fact that the current I through S will be positive or negative, depending on the assumed positive sense for l. The reversal of the positive sense for l, for example, will reverse that for the outward sense for S and thus the sign of the current I. However, Ampère's law in (25-12) must be, and is, independent of this choice; it is valid for either one.

To help fix these ideas, consider in Figure 25-19a a current i, say in a thin wire, going through a surface S in the direction shown. For the given sense of the bounding curve l, the current through S is positive. Hence Ampère's law in this case becomes

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0 i$$

On the other hand, if the sense of the curve relative to the direction of the current flow is as shown in Figure 25-19b, the corresponding relation is

$$\int_l \mathbf{B} \cdot d\mathbf{l} = -\mu_0 i$$

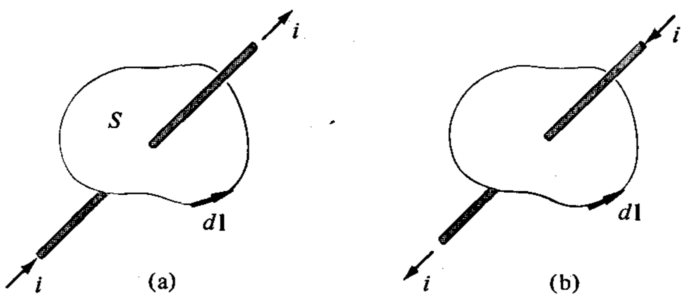


Figure 25-19

In a similar way, since the total current passing through the surface S in Figure 25-20a is $(i_1 + i_2)$, Ampère's law here assumes the form

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0(i_1 + i_2)$$

Correspondingly, for the currents shown in Figure 25-20b the analogous

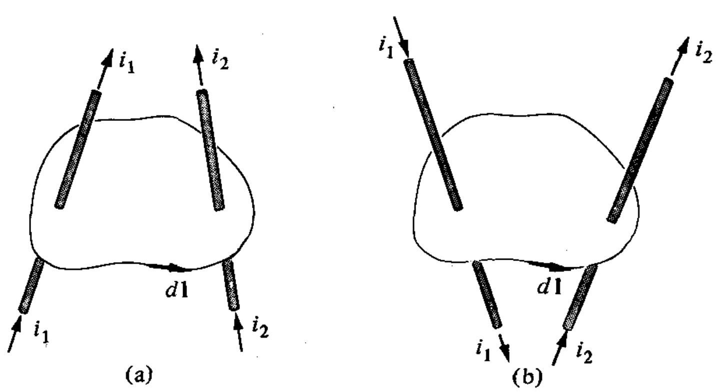


Figure 25-20

formula is

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0(i_2 - i_1)$$

Example 25-10 A uniform current i flows in a cylinder of radius a, and an identical current i flows in a nearby thin wire (see Figure 25-21). Evaluate, by use of Ampère's law, the line integral $\oint \mathbf{B} \cdot d\mathbf{l}$ about each of the paths A, B, C, and D, with the sense of direction specified in the figure.

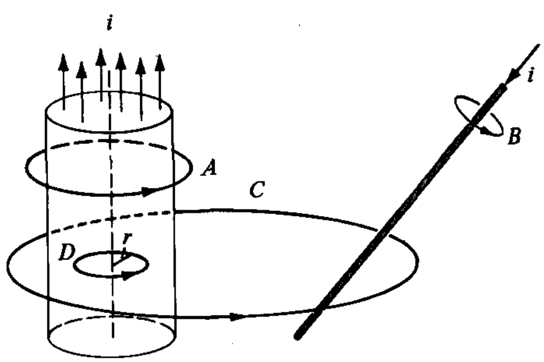


Figure 25-21

Solution The total current through path A is i, and hence

$$\oint_A \mathbf{B} \cdot d\mathbf{l} = \mu_0 i$$

Similarly, through path B the total current is i, but this time the current is directed opposite to that associated with the assumed sense of the path. Thus we have

$$\oint_B \mathbf{B} \cdot d\mathbf{l} = -\mu_0 i$$

The total current through path C consists of an upward flow i through the cylinder plus a downward flow i through the wire. Accordingly the net current through path C vanishes, and thus

$$\oint_C \mathbf{B} \cdot d\mathbf{l} = 0$$

Finally, since the current through the cylinder is uniform, it follows that the current density is $i/\pi a^2$. This implies that the total current through path D (the circle of radius $r < a$ in the figure) is $i\pi r^2/\pi a^2$. Hence, by Ampère's law,

$$\oint_D \mathbf{B} \cdot d\mathbf{l} = \frac{\mu_0 i r^2}{a^2}$$

It should be emphasized that these formulas are valid even in the presence of other nearby currents, provided only that these currents do not flow through any surfaces bounded by these four integration paths.

25-10 Applications of Ampère's law

We have seen previously that by use of the Biot-Savart law it is possible to calculate, in principle, the magnetic induction associated with *any* given distribution of currents. The purpose of this section is to show how these calculations may be simplified in certain cases by use of Ampère's law. Just as for the analogous calculations of the electrostatic field by use of Gauss' law, this method is applicable only for highly symmetric current distributions.

As a first application consider again, in Figure 25-12, the **B**-field at a perpendicular distance r from a very long wire carrying a current i. Let us apply Ampère's law to this case by selecting a circular path of radius r and with the center on the wire. Assuming that the **B**-field lines are circles concentric with the wire, we find by use of (25-12) that

$$\mu_0 i = \oint_l \mathbf{B} \cdot d\mathbf{l} = B \oint dl = 2\pi r B$$

where the second equality follows since $\mathbf{B}$ and $d\mathbf{l}$ are parallel, so that $\mathbf{B} \cdot d\mathbf{l} = B\,dl$, and from the fact that B is constant along the path of integration and may thus be taken out from under the integral. Solving the last equality for B we regain (25-6). The validity of Ampère's law is thus confirmed for this case.

As a second application of Ampère's law, consider the magnetic induction in the *interior* of an infinitely long cylinder of radius a, through which flows a uniform axial current density $\mathbf{j}$ corresponding to a current $i = j\pi a^2$. Just as for the external field, the lines of magnetic induction in the interior of the wire must be circles concentric with the axis of the wire. For if there were, say, a radial component to $\mathbf{B}$ at any point inside, then either the symmetry or the continuity property of the $\mathbf{B}$ lines would be violated. Further, the strength of $\mathbf{B}$ can vary only with the distance r from the axis, so $\mathbf{B}$ must have the same magnitude at all points of any circle concentric with the axis.

To calculate B let us apply Ampère's law to a circle of radius r ($<a$) centered on the axis of the wire, as shown in Figure 25-22. The total current

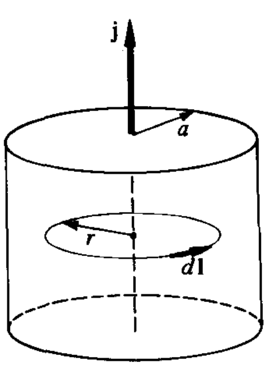


Figure 25-22

through this circular path is

$$\int_S \mathbf{j} \cdot d\mathbf{S} = j \int_S dS = \pi r^2 j = \frac{r^2}{a^2} i$$

where S has been taken to be the disk that is bounded by the circular path. Substitution into Ampère's law in (25-12) leads to

$$\frac{\mu_0 i r^2}{a^2} = \oint \mathbf{B} \cdot d\mathbf{l} = B \oint dl = 2\pi r B$$

since the field lines are assumed to be circles concentric with the axis. Solving for B we obtain

$$B = \frac{\mu_0 i r}{2\pi a^2} \qquad (r \leq a) \tag{25-13}$$

so B vanishes along the axis of the cylinder and rises linearly with radius to assume its maximum value $\mu_0 i/2\pi a$ on the surface. Outside the wire, B decreases for increasing values for r in accordance with (25-6). The field lines both inside and outside are circles concentric with the cylinder axis, and with their sense given by the right-hand rule. Figure 25-23 shows a plot of B as a function of r both inside and outside the wire.

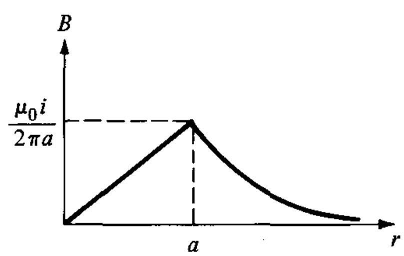


Figure 25-23

As a third and final application of Ampère's law we now show that the relation

$$B = \mu_0 n i \tag{25-9}$$

which characterizes the **B**-field on the axis of an infinitely long solenoid is valid *throughout the interior* of what is known as an *ideal solenoid.* A solenoid is said to be ideal if it is very long and is sufficiently tightly wound that the current flow on its surface may be thought of as a continuous sheet of current. Strictly speaking, no solenoid is truly ideal. However, just as for the parallel-plate capacitor—for which we neglected the fringing field at the edges of the plates—if we agree to neglect the leakage of **B**-lines out of the coil, then we can consider it to be ideal.

Figure 25-24 shows the field lines associated with the current in an ordinary solenoid. Note that there is an outward "leakage" of field lines between the wires, and that in the region near the center the field lines tend to be parallel to the axis. It is plausible to expect that the longer is the

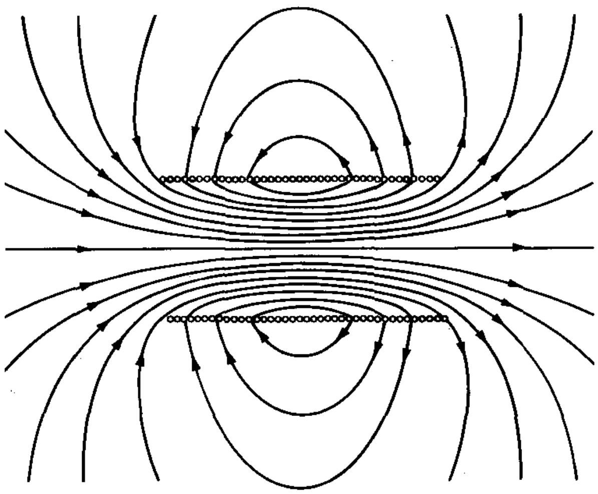


Figure 25-24

solenoid and the closer together are the turns, the more extended will be this region in which the field lines are parallel to the axis. The *ideal solenoid* is then the limiting form for which the field outside vanishes, while inside the field lines are everywhere parallel to the axis.

To demonstrate the validity of (25-9) for all points inside an ideal solenoid, let us apply Ampère's law to the rectangular path *abcd* in Figure 25-25 with the segment $\overline{ab}$ on the axis of the coil. Since the entire path lies inside the coil, no current passes through it. Hence, by Ampère's law,

$$0 = \oint \mathbf{B} \cdot d\mathbf{l} = \int_a^b \mathbf{B} \cdot d\mathbf{l} + \int_b^c \mathbf{B} \cdot d\mathbf{l} + \int_c^d \mathbf{B} \cdot d\mathbf{l} + \int_d^a \mathbf{B} \cdot d\mathbf{l}$$

$$= \int_a^b \mathbf{B} \cdot d\mathbf{l} + \int_c^d \mathbf{B} \cdot d\mathbf{l}$$

where the third equality follows since by hypothesis the field lines are parallel to the axis and hence perpendicular to the paths $\overline{bc}$ and $\overline{da}$. Now since the path $\overline{ab}$ lies along the axis of the coil, and since there the field has according to (25-9) the value $\mu_0 ni$, it follows that

$$\int_a^b \mathbf{B} \cdot d\mathbf{l} = \mu_0 ni \, \Delta h$$

where Δh is the distance from a to b. In a similar way, if B is the value of

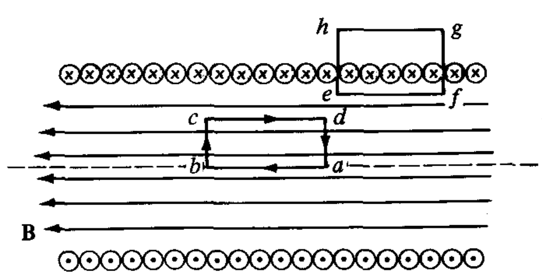


Figure 25-25

the field along the path $\overline{cd}$, then

$$\int_c^d \mathbf{B} \cdot d\mathbf{l} = -B\,\Delta h$$

where the minus sign reflects the fact that $\mathbf{B}$ and $d\mathbf{l}$ are antiparallel here. Combining the last three relations and canceling the common factor Δh, we find that

$$B = \mu_0 n i \tag{25-14}$$

Finally, since this argument can be repeated *everywhere* in the interior of the solenoid, it follows that the field in the interior has everywhere the same value as on the axis.

In the same way, but making use of the rectangular path *efgh*, with $\overline{gh}$ *outside* of the coil, it is left as an exercise to show that the **B**-field outside of an ideal solenoid vanishes.

Example 25-11 A uniform current of density j flows along (down into the plane of Figure 25-26) an infinitely long and conducting cylinder of inner radius a and outer radius b. Calculate the **B**-field everywhere.

Solution Because of the fact that the cylinder is infinitely long, it follows, just as above, that the field lines are circles concentric with the axis. Accordingly, let us apply Ampère's law successively to the three regions: $r > b$; $b \geq r \geq a$; and $r < a$, by use in each case of a circle of appropriate radius. In the region $r > b$ there results

$$\oint \mathbf{B} \cdot d\mathbf{l} = 2\pi r B = \mu_0 j \pi (b^2 - a^2)$$

where the second equality follows since the total current through the circle of radius r ($>b$) is $j\pi(b^2 - a^2)$. Hence

$$B = \frac{\mu_0 j (b^2 - a^2)}{2r} \qquad (r > b)$$

In a similar way, for the region $b \geq r \geq a$,

$$B = \frac{\mu_0 j (r^2 - a^2)}{2r} \qquad (b \geq r \geq a)$$

Finally, for the region $r \leq a$,

$$B = 0 \qquad (r < a)$$

since no current flows through a circle of radius $r < a$.

Example 25-12 A *toroid* is a solenoid that has been bent into the form of a *torus*. See Figure 25-27. Assuming that the N turns of the toroid are very close together, so that the field lines are confined to its interior, calculate **B** at a distance r ($a < r < b$) from the center of the toroid. See Figure 25-27b.

Solution Using symmetry arguments, it follows that the **B**-field lines inside the torus must be circles concentric with the center of the torus. Their sense is

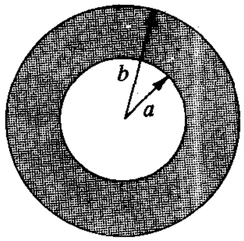


Figure 25-26

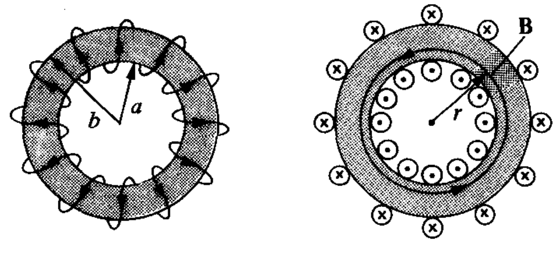


Figure 25-27

determined by the right-hand rule, and for the current flow in the figure they are directed as shown.

To calculate B, let us apply Ampère's law to a circle of radius r inside the torus. Since there are N turns about the torus, it follows that the total current through this path is Ni. (Note that for values of r for which this path is *outside* the torus, the associated current flow vanishes.) Assuming, as is implied in the figure, that $\mathbf{B}$ depends only on r and is tangent to this circular path, we find that

$$\mu_0 Ni = \oint \mathbf{B} \cdot d\mathbf{l} = B \oint dl = 2\pi r B$$

and this leads to

$$B = \frac{\mu_0 Ni}{2\pi r} \tag{25-15}$$

Unlike the ideal solenoid, inside of which the field is constant, in the interior of a toroid, $\mathbf{B}$ varies with radius in accordance with this formula. In both cases, $\mathbf{B}$ vanishes outside.

25-11 Summary of important formulas

The force $\mathbf{F}$ on a particle of charge q traveling at the velocity $\mathbf{v}$ through a magnetic induction field $\mathbf{B}$ is

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} \tag{25-1}$$

Experiment shows that the field $\mathbf{B}$ is independent of q and $\mathbf{v}$ and thus is determined exclusively by the magnetic sources. The properties of the particle enter *only* through the factor $q\mathbf{v}$.

The magnetic induction $d\mathbf{B}$ due to an infinitesimal current element $i\,d\mathbf{l}$ in a thin wire is given by the Biot-Savart law

$$d\mathbf{B} = \frac{\mu_0 i}{4\pi} \frac{d\mathbf{l} \times \mathbf{r}}{r^3} \tag{25-3}$$

with $\mathbf{r}$ the position vector from $d\mathbf{l}$ to the field point. The total field produced by a collection of currents is found by integrating (25-3) around all closed current paths.

Gauss' law for magnetism states that the magnetic flux out of every closed surface vanishes:

$$\oint_S \mathbf{B} \cdot d\mathbf{S} = 0 \tag{25-11}$$

where the integral is to be carried out over the arbitrary, but closed, surface S. Ampère's law states that for l an arbitrary, closed curve,

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0 I \tag{25-12}$$

where I is the net current flowing along the outward sense of an arbitrary *open* surface bounded by l. See Figure 25-18.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** north pole; **(b)** magnetic monopole; **(c)** Gauss' law for magnetism; **(d)** magnetic flux; and **(e)** Ampère's law.
2. Consider the earth as a huge magnet. In view of the fact that the "north pole" of a magnetized needle points north, is the northernmost magnetic pole of the earth a north or a south magnetic pole? Explain.
3. Suppose that the magnetic induction field of the earth were produced by circular currents flowing in its equatorial plane. Would these currents flow eastward or westward around the earth?
4. If a magnetized needle were free to rotate simultaneously in both a vertical and a horizontal plane, which way would the needle be pointing near the north pole of the earth? Near the south pole?
5. An electron moving at some fixed velocity enters a region of space where there is a uniform **B**-field. Explain how it is possible for the electron to experience *no* force under these circumstances.
6. Consider the three vectors **F**, **v**, and **B** in (25-1). Which two pairs of these three are always perpendicular to each other?
7. A proton traveling at a velocity of 5.0×10^5 m/s enters a region in which there is a time-independent magnetic induction **B**. What must be its speed when it leaves this region? Can anything be said in general about its final direction of motion? Explain.
8. Explain why (25-1) suffices to define the **B**-field for a given magnetic source even though the component of **B** along **v** does not enter this formula.
9. A particle of charge q is at rest in a region of space where there exists a **B**-field produced by nearby currents. What is the force on the particle, assuming that there is no electric field present?
10. Consider again the situation of Question 9, but this time from the viewpoint of a moving observer with respect to whom the particle has the velocity **v**. What, according to this observer, is the force on the particle? Explain your answer in light of (25-1) and the fact that there were no electric fields present originally.
11. Contrast and compare the two right-hand rules, which give the directions of **B** associated, respectively, with the current in a straight wire and that in a solenoid. Review the arguments used for deriving these two rules.
12. An observer sees current flowing counterclockwise about a loop of

wire shaped in the form of a triangle. Are the field lines in the plane of the loop directed toward or away from this observer? Justify your answer.

13. An explorer is looking at his compass when suddenly a horizontal bolt of lightning flashes overhead. Assuming that the associated current is directed due north, will the compass needle be deflected eastward or westward? Explain.
14. Devise a circuit analogous to that in Figure 25-5 so that the currents in the two hanging wires are in the same direction.
15. In view of result in (25-13) is it correct to say that the magnetic induction at a distance r from the axis of a uniform current is due only to those currents flowing inside the cylinder of radius r? Explain.
16. Consider the flow of a uniform current along the axis of a cylinder that has an elliptical cross section. Why can we *not* calculate the magnetic induction by use only of Ampère's law? Explain.
17. In view of the fact that parallel currents attract, do you expect the neighboring turns of wire in a solenoid to attract or to repel each other?
18. A particle is suspended from the lower end of a spring, whose axis is vertical. If a steady current is somehow sent through the spring, will the new equilibrium position of the particle be above or below its original one? Does it matter in which direction current flows?
19. A proton is traveling parallel to a straight wire along which flows a current i. If the proton is moving in the direction of the current, what is the direction of the force on the proton?
20. Suppose that a current of $(20/4\pi)$ amperes flows around the rectangular loop in Figure 25-28. Evaluate $\oint \mathbf{B} \cdot d\mathbf{l}$ for each of the four paths A, B, C, and D.

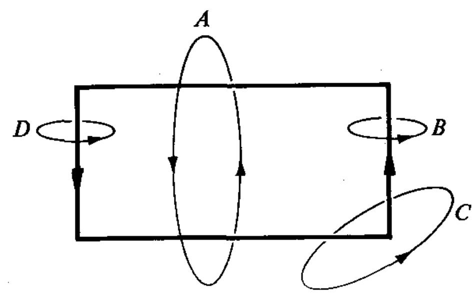


Figure 25-28

21. Show by use of Ampère's law that there does not exist a scalar potential function whose derivatives are the components of the **B**-field. (*Hint*: Assume the existence of such a function and show that it leads to a contradiction.)
22. Devise an experiment to confirm *Gauss' law* for magnetism.
23. Suppose you have available an apparatus to measure the magnetic induction at any point on earth. Devise a *thought experiment* by means of which you could ascertain whether the magnetic induction of the earth is due to currents flowing in its interior or to currents in the space outside. (*Hint*: Use the apparatus in conjunction with Ampère's law.)

PROBLEMS

1. A particle having a charge of 2.0 μC is traveling at a velocity of 3.0×10^5 m/s in a region of space where there exists a magnetic induction of strength 2.5×10^{-3} tesla directed due north. Calculate the magnitude and direction of the force on this particle at an instant when it is traveling: **(a)** vertically upward; **(b)** due south; **(c)** in a northeasterly direction.
2. Calculate the magnitude of the

force on a proton traveling through the upper atmosphere, where the earth's field has the value 4.0×10^{-5} tesla. Assume the proton to be traveling in a direction making an angle 45° with respect to **B** and at a speed 6.0×10^{5} m/s.

3. A cathode-ray tube is oriented so that the electron steam is directed vertically upward. Assuming that the electrons have a velocity of 10^{7} m/s, and that the horizontal component of the earth's field has a strength of 7.0×10^{-5} tesla, calculate **(a)** the strength and direction of the magnetic force on each electron and **(b)** the acceleration of an electron due to this force.

4. A proton beam is sent into a region of space where there exists a uniform magnetic induction. In terms of a Cartesian coordinate system set up in this region, it is observed that the beam is undeflected if it travels along the positive z-axis, but if it is sent along the direction of the positive x-axis, then it is deflected along the direction of the positive y-axis. Based on these data, what is the direction of the magnetic induction?

5. Repeat Problem 4, but assume that the observed results are for an electron beam.

6. A proton makes circular orbits in a cyclotron with a speed of 3.0×10^{7} m/s. If its acceleration is observed to be 1.2×10^{15} m/s^2, calculate **(a)** the force on the proton and **(b)** the strength of the **B**-field perpendicular to the plane of the orbit.

7. In a certain Cartesian coordinate system a current element $i\,d\mathbf{l}$ is located at the origin and oriented along the positive z-axis. Calculate the direction and magnitude of **B** produced by this element at the following points (assume that the parameter a is in each case a positive distance): **(a)** $(0, 0, a)$; **(b)** $(0, a, 0)$; **(c)** $(-a, 0, 0)$; and **(d)** $(a, a, 0)$.

8. A current of 5.0 amperes flows along a very long and straight wire. Calculate the strength of the **B**-field at two points, which are at perpendicular distances of 1.0×10^{-3} meter and 2.0 meters from the wire, respectively. Under what circumstance will the fields at these two points be parallel? Under what circumstance will they be antiparallel?

***9.** Show, by use of the Biot-Savart law, that the **B**-field at a distance **r** from a particle of charge q that is traveling at a velcoity **v** is

$$\mathbf{B} = \frac{\mu_0}{4\pi} q \frac{\mathbf{v} \times \mathbf{r}}{r^3}$$

10. Consider an electron orbiting about a proton in a hydrogen atom in a circle of radius 0.53×10^{-10} meter with a speed 2.2×10^{6} m/s. Making use of the result of Problem 9, calculate the strength of the **B**-field at the position of the proton due to the electron's motion.

11. If the wire in Figure 25-15 had a total length $2l$ and the point P is on the perpendicular bisector with coordinates $(x, 0, 0)$, show that **B** is directed along the y-axis in the figure and has the magnitude

$$B = \frac{\mu_0 i}{2\pi x} \frac{l}{[x^2 + l^2]^{1/2}}$$

(*Hint*: Use the methods of Example 25-3, but restrict the region of integration to the interval $-l \leq z \leq l$.)

12. A current i flows about a square loop of side a. By use of the result of Problem 11 show that B at the center of the loop is given by

$$B = \frac{2\sqrt{2}\mu_0 i}{\pi a}$$

and that its direction is given by the right-hand rule for a circular loop.

13. A 10-ampere current flows around a

loop of wire in the shape of an equilateral triangle 50 cm on a side. By use of the result of Problem 11, calculate B at the center of the loop.

14. Repeat Problem 12, but this time calculate B at a point P a distance b above the plane of the square loop and along the perpendicular through its center. (*Hint*: Recall that **B** is a vector, and vector addition must be used to add together the contributions from each side of the square.)

*15. A wire of length $2l$ carries a current i and lies along the z-axis of a certain coordinate system with its center at the origin. Show that the **B**-field at a point with the coordinates $(x, 0, z)$ has the magnitude

$$B = \frac{\mu_0 i}{4\pi x}\left\{\frac{l-z}{[(l-z)^2+x^2]^{1/2}} + \frac{l+z}{[(l+z)^2+x^2]^{1/2}}\right\}$$

and determine its direction. Show also that this formula reduces to the result in Problem 11 for an appropriate choice of variables.

16. Two infinite parallel wires are separated by a distance $2a$ and carry currents i in opposite directions, as shown in Figure 25-29. Calculate **B** at a point P, which lies at a distance b along the perpendicular bisector, so that in terms of the coordinate system shown, the coordinates of P are (b, a). What should your answer reduce to when $b = 0$?

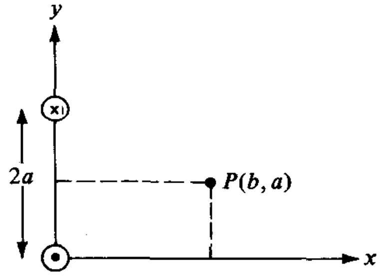


Figure 25-29

*17. Repeat Problem 16, but calculate the field at a general point with coordinates (x, y). Is it an essential restriction that we do not introduce the z-coordinate of the field point?

18. Repeat Problem 17, but assume this time that the currents in the two wires are *both* directed perpendicularly down into the plane of Figure 25-29.

19. What current must flow around a circular loop of wire of radius 50 cm so that the field at the center is 5×10^{-5} tesla? What current must flow if the field is to have a strength of 2 teslas?

20. Show, by use of the result of Example 25-8, that if $\partial B/\partial z = 0$ and $\partial^2 B/\partial z^2 = 0$, then $z = c = a/2$. Explain why you might expect the resultant *Helmholtz coil* to have a relatively constant magnetic induction along the axis of, and midway between, the two loops.

21. A current i flows in a segment of a circular loop of radius a and angle α, as shown in Figure 25-30. Calculate **B** at the center C of the loop, disregarding the wires that feed the current.

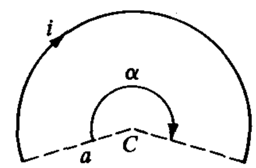


Figure 25-30

22. Suppose that a circular current circled the earth at the equator. What would the strength of this current have to be to produce the observed field at the poles of about 7.5×10^{-5} tesla? Would the current flow from east to west, or in the opposite direction?

23. Calculate at the point P the strength of the magnetic induction due to the current i through the wire in Figure 25-31. (*Hint*: Use the result of Problem 21.)

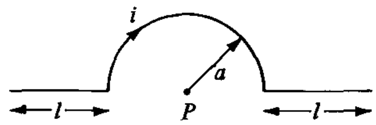


Figure 25-31

24. A circular *line charge* of radius a and charge per unit length λ rotates at an angular velocity ω about its axis.

(a) Show that this motion corresponds to a current i flowing in a circular loop of radius a and given by

$$i = \lambda a \omega$$

(b) Calculate $\mathbf{B}$ at a point on the axis of the line charge and at a distance b from its center.

***25.** A disk of radius a carries a uniform charge per unit area σ and rotates with an angular velocity ω about its axis. See Figure 25-32.

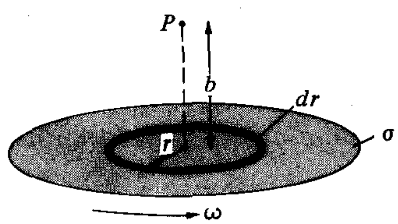


Figure 25-32

(a) Show that the magnetic induction dB at point P on the axis due to a circular ring of radius r and thickness dr is

$$dB = \frac{\mu_0 \sigma \omega r^3\, dr}{2[r^2 + b^2]^{3/2}}$$

(*Hint*: Make use of the results of Problem 24.)

(b) Show that the total field B at the point P is

$$B = \frac{\mu_0 \sigma \omega}{2}\left[\frac{a^2 + 2b^2}{(a^2 + b^2)^{1/2}} - 2b\right]$$

26. A solenoid of radius 0.5 cm and length 20 cm carries a current of 10 amperes and has 1000 turns. Calculate the strength of the magnetic induction on the axis of the coil at the following points: (a) the center of the coil; (b) the edge of the coil; (c) a distance of 100 cm from the center of the coil.

27. For the solenoid described in Problem 26 make a plot of B as a function of position on the axis of the coil.

28. How many turns per unit length are required so that a very long solenoid which carries a current of 2.0 amperes will have a magnetic induction on its axis of 1.5×10^{-2} tesla?

29. A current of 0.5 ampere flows around a solenoid of radius 1.0 cm and of length 40 cm. If the strength of the uniform magnetic induction near the center of the solenoid is 10^{-3} tesla, how many turns per unit length are there in the solenoid?

30. Suppose that the field on the axis of a very long solenoid having 1000 turns per meter is 5.0×10^{-3} tesla. (a) What is the current? (b) If now a wire that carries a current of 10 amperes is wrapped with n turns per unit length around the original solenoid in such a way that the field on the axis is reduced to 2.5×10^{-3} tesla, calculate n. Are the current flows parallel in these two coils?

31. A proton is traveling at a speed of 5.0×10^5 m/s on the axis of a solenoid, which has 1000 turns per meter and carries a current of 2.0 amperes. Calculate the acceleration of the proton. In what way would your answer differ if it were moving parallel to but *not* on the axis.

32. A man walks due north underneath and parallel to a power line in which there flows a direct current of 100 amperes. If he is 10 meters below the line, what magnetic induction—beyond that due to the earth—would he measure? Would this seriously interfere with a compass reading at this point?

***33.** Figure 25-33 shows a cut through an infinite array of n (per unit length) parallel wires, each carrying a current i directed as shown.

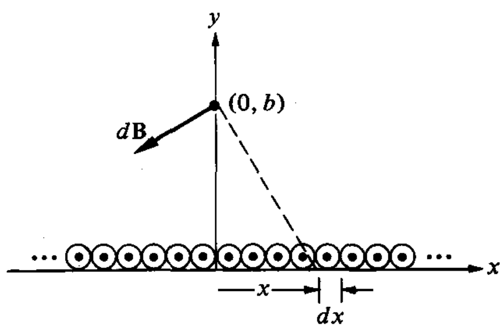


Figure 25-33

(a) Show that the field $d\mathbf{B}$ at the point with coordinates (0, b) due to the $n\,dx$ wires located between x and $(x + dx)$ is

$$d\mathbf{B} = -\frac{\mu_0 ni}{2\pi(b^2 + x^2)}(\mathbf{i}b + \mathbf{j}x)\,dx$$

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors along the x- and y-axes, respectively.

(b) By integration show that the field for $y > 0$ due to all of the currents is uniform—that is, independent of b—and is given by

$$\mathbf{B} = -\frac{1}{2}\mu_0 ni\mathbf{i} \qquad y > 0$$

(c) Show in a similar way that for points below the currents, that is, for $y < 0$,

$$\mathbf{B} = \frac{1}{2}\mu_0 ni\mathbf{i} \qquad y < 0$$

The field lines for this current distribution are shown in Figure 25-34.

34. Confirm the validity of the results for (b) and (c) of Problem 33 by applying Ampère's law to the rectangular path *abcd* in Figure 25-34.

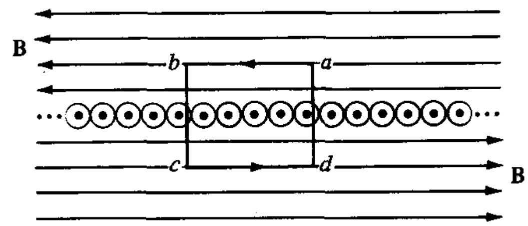


Figure 25-34

35. Suppose that the cylinder in Figure 25-35 represents a solenoid of N turns and length l, around which flows a current i so that the field on the axis of the solenoid is directed upward. Evaluate $\oint \mathbf{B} \cdot d\mathbf{l}$ for the paths **(a)** *ABCDA*; **(b)** *abcda*; **(c)** l_1; and **(d)** l_2.

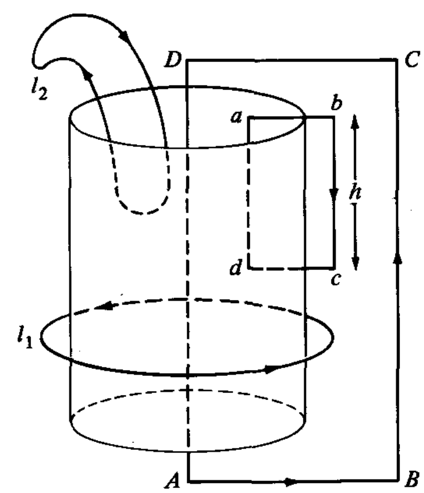


Figure 25-35

***36.** Show that it is impossible for a *uniform* magnetic induction to drop abruptly to zero in a current-free region. (*Hint*: Suppose, as in Figure 25-36, that there is a uniform field in the half-space $x < 0$, and that B is zero for $x > 0$. By applying Ampère's law to the path *abcd* show that this leads to a contradiction. What conclusions can you draw about the magnetic induction based on this result?)

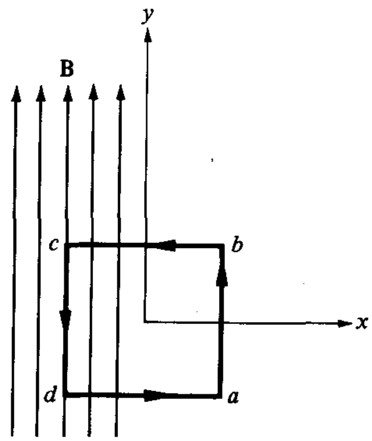


Figure 25-36

***37.** Consider, in Figure 25-37, a circular coil of wire of radius a and carrying a current i.

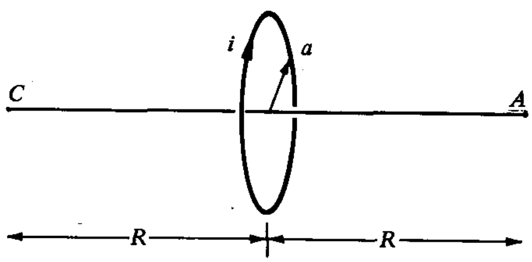


Figure 25-37

(a) Evaluate the line integral

$$\mathcal{J} = \int_A^C \mathbf{B} \cdot d\mathbf{l}$$

along the axis of the loop a symmetric distance of $2R$.

(b) Show that in the limit as $R \to \infty$,

$$\mathcal{J} = \mu_0 i$$

(c) Justify your result to (b) by use of Ampère's law. (*Hint*: Close the integration path AC by a large semicircle and show that $\int \mathbf{B} \cdot d\mathbf{l}$ along this path vanishes as $R \to \infty$.)

***38.** Repeat Problem 37, but this time along the axis of a solenoid of length l and N turns.

***39.** A very large coaxial cable consists, as in Figure 25-38, of an inner solid conductor of radius a and an outer one of inner radius b and outer radius c. Assuming equal and opposite uniform currents i in these conductors, calculate the magnetic induction for **(a)** $r \leq a$; **(b)** $a \leq r \leq b$; **(c)** $b \leq r \leq c$; and **(d)** $r \geq c$.

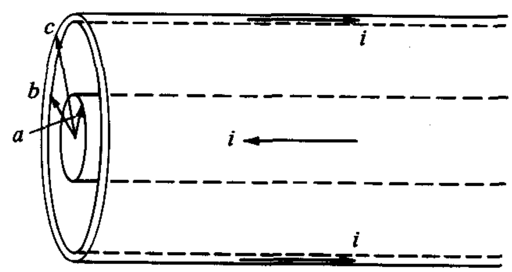


Figure 25-38

26 Magnetic forces

The only justification for our concepts and systems of concepts is that they serve to represent the complex of our experiences; beyond this, they have no legitimacy.

ALBERT EINSTEIN (1879–1955)

26-1 Introduction

In Chapter 25 the magnetic induction field was defined in terms of the force acting on a moving charged particle. The focus there, however, was not on this force itself, but rather on the problem of calculating the fields associated with certain idealized current distributions. The purpose of this chapter is to present a more detailed analysis of this force and to consider it as it affects both the motions of individual charged particles as well as macroscopic current flows. The principles underlying the operation of a number of devices including cyclotrons, mass spectrometers, and galvanometers are, as we shall see, readily explained in terms of this force.

26-2 The motion of a charged particle in a uniform B-field

Consider a particle of charge q and mass m moving in a uniform magnetic induction $\mathbf{B}$. According to (25-1) the force on it is $q\mathbf{v} \times \mathbf{B}$ and thus it follows,

from Newton's second law, that the equation of motion for the particle is

$$m\frac{d\mathbf{v}}{dt} = q\mathbf{v}\times\mathbf{B} \tag{26-1}$$

with $\mathbf{v}$ its instantaneous velocity. All the details of the particle's trajectory can be deduced from this relation by solving it subject to an appropriate set of initial conditions. However, rather than proceeding in this way it is instructive to solve it indirectly by use of a much simpler heuristic approach which is available in this particular case.

Suppose, first, that initially the velocity of the particle is perpendicular to the uniform **B**-field. Specifically, let us assume that the **B**-field lines are directed perpendicularly down into the plane of Figure 26-1 and, following convention, let us represent these lines by an array of crosses. (Correspondingly, an array of dots is normally used to represent **B**-field lines emerging from the plane of a figure toward the reader.) The force **F** on the particle is directed as shown in Figure 26-1a for $q > 0$ and in Figure 26-1b for $q < 0$. In both cases **F** is perpendicular to **v** throughout the motion. Hence, only the direction, but not the magnitude, of the particle's velocity can change during its subsequent motion (see Example 25-2). It follows that the orbit of the particle must be a circle whose plane is perpendicular to **B**. Further, the magnitude of the velocity of the particle along this orbit must be constant and equal to its initial value $v = |\mathbf{v}|$.

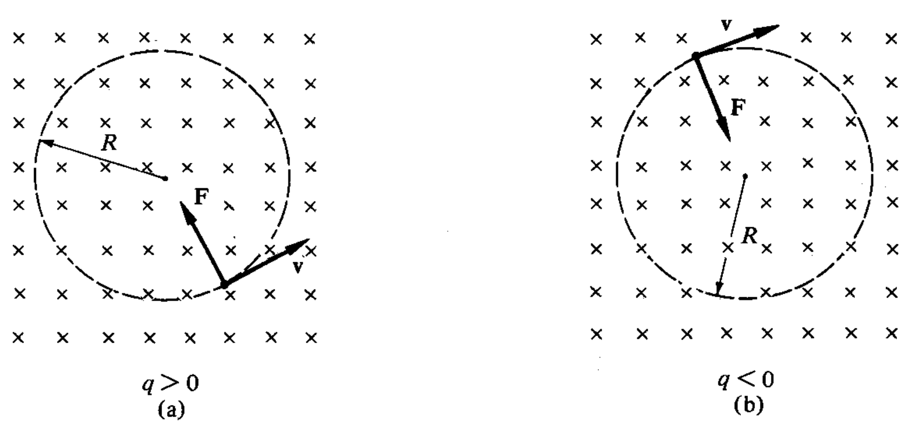


Figure 26-1

Now according to the analysis of Section 3-10, a particle such as this which travels with a uniform speed v around a circle of radius R undergoes a *centripetal* acceleration $\mathbf{a}_c$, which is directed radially inward and has the magnitude v^2/R. Reference to Figure 26-1 shows, then, that for a uniform field we may express (26-1) in the form

$$qvB = m\frac{v^2}{R} \tag{26-2}$$

and this is the fundamental relation from which all orbital parameters can be

determined. For example, solving (26-2) for the radius we obtain

$$R = \frac{mv}{qB} \tag{26-3}$$

so that R is directly proportional to the speed, v, of the particle and with the coefficient of proportionality the known ratio m/qB. The angular velocity of a particle going with a uniform speed, v, around a circle of radius R is v/R. Hence, solving (26-3) for this ratio, we obtain

$$\omega_c = \frac{v}{R} = \frac{qB}{m} \tag{26-4}$$

where the quantity ω_c is the *cyclotron frequency* associated with this orbit. A most striking feature of this formula for ω_c is that it depends *only* on B and on the ratio q/m and is thus independent of the velocity of the particle. This means that a collection of *identical* charged particles will orbit, in a given **B**-field, with the same angular velocity ω_c regardless of their initial velocities. According to (26-3), however, the radii of these orbits will, in general, be different. The period P of this motion is

$$P = \frac{2\pi}{\omega_c} = \frac{2\pi m}{qB} \tag{26-5}$$

If the initial velocity **v** of a particle is *not* perpendicular to **B**, then the orbit of the particle is not confined to a plane perpendicular to **B** as in Figure 26-1. To treat this more complex case, suppose that $v_\perp$ is the component of **v** perpendicular to **B** and $v_\parallel$ the corresponding component along **B**. As far as the motion at right angles to **B** is concerned, all of the relations in (26-2) through (26-5) are still valid provided that v is everywhere replaced by $v_\perp$. With regard to $v_\parallel$, note from (26-1) and the definition of a cross product that there is no component of the force **F** along the direction of **B**. Hence the particle will continue to move with its initial velocity $v_\parallel$ along this direction. The motion of the particle thus consists of two parts: (1) a motion at the initial velocity $v_\parallel$ along the direction of **B**; and (2) a motion in a circular orbit at the uniform speed $v_\perp$ at right angles to **B**. When viewed by an external observer, the particle appears to move in a *helix*, as shown in Figure 26-2.

Figure 26-3 illustrates the circular paths followed by charged particles in a uniform **B**-field. Shown are the actual tracks produced in the collision of a 300 GeV (3.0×10^{11} eV) proton with a stationary one in the 30-in. hydrogen bubble chamber at the Fermi National Accelerator laboratory. Twenty-six charged particles are produced. Assuming that the inward spiraling tracks are due to electrons, is the **B**-field directed into or out of the plane of the photograph?

Example 26-1 A proton is moving at right angles to a uniform magnetic induction having a strength of 2.0 tesla.

(a) Calculate the cyclotron frequency.

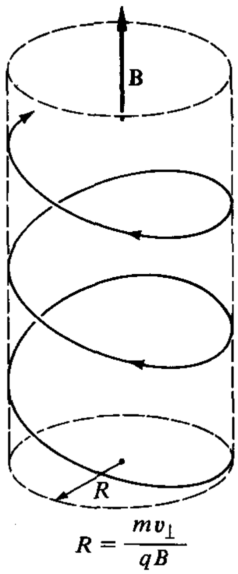


Figure 26-2

(b) How long does it take the proton to make one full turn around its orbit?
(c) If the radius of the orbit is 5.0 cm, what is the velocity of the proton?
(d) What is the proton's energy?

Solution

(a) Using the fact that for a proton $m = 1.67 \times 10^{-27}$ kg and $q = 1.6 \times 10^{-19}$ coulomb, we find, by use of (26-4), that

$$\omega_c = \frac{qB}{m} = \frac{(1.6 \times 10^{-19}\text{ C}) \times (2.0\text{ T})}{1.67 \times 10^{-27}\text{ kg}}$$
$$= 1.9 \times 10^{8}\text{ rad/s}$$

(b) Making use of this result we find for the period P of the motion

$$P = \frac{2\pi}{\omega_c} = \frac{2\pi}{1.9 \times 10^{8}\text{ rad/s}} = 3.3 \times 10^{-8}\text{ s}$$

In other words, the proton makes one complete revolution about its circular orbit in 3.3×10^{-8} second.

(c) Solving (26-3) for v we find

$$v = R\frac{qB}{m} = R\omega_c = (0.05\text{ m}) \times (1.9 \times 10^{8}\text{ rad/s})$$
$$= 9.5 \times 10^{6}\text{ m/s}$$

(d) The energy E of the proton is all kinetic. Hence

$$E = \tfrac{1}{2}mv^2 = \tfrac{1}{2} \times 1.67 \times 10^{-27}\text{ kg} \times (9.5 \times 10^{6}\text{ m/s})^2$$
$$= 7.5 \times 10^{-14}\text{ J}$$

In connection with elementary particles it is customary to express energies in units of the "MeV." This is defined to be the energy gained by a proton in dropping through a

Figure 26-3 A proton with an energy of 300 GeV producing 26 charged particles in a 30-in. hydrogen bubble chamber. (Courtesy Fermi National Accelerator Laboratory.)

potential difference of 10^6 volts. It has the approximate value

$$1 \text{ MeV} = 1.6 \times 10^{-13} \text{ J}$$

In terms of this unit, E has the value

$$E = 7.5 \times 10^{-14} \text{ J} \times \left(\frac{1 \text{ MeV}}{1.6 \times 10^{-13} \text{ J}}\right)$$
$$= 0.47 \text{ MeV}$$

Example 26-2 Consider in Figure 26-4 a collimated beam of particles of various charges, masses, and velocities entering via slit A a region of space in which there is a uniform electric field **E** directed vertically downward and a uniform magnetic

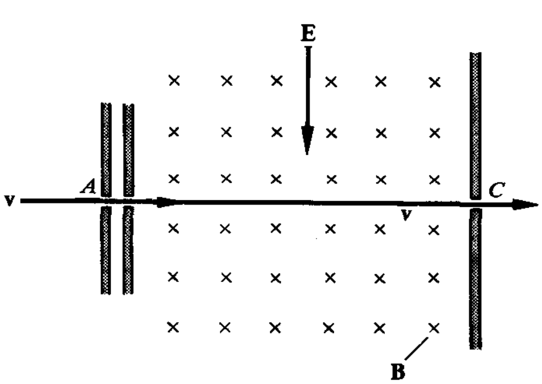


Figure 26-4

induction perpendicular to **E**. What can be concluded about the velocity, the charge, and the mass of the particles that eventually emerge at slit C?

Solution The subset of particles that arrive at C consists of those that continue to travel in a straight line and thus experience no force on traversing this region. All others will be accelerated at right angles to this direction. Hence if **v** is the velocity of any particle which reaches C, then

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = 0$$

where q is its charge. Since the vectors **E**, **B**, and **v** are, as shown in Figure 26-4, mutually orthogonal it follows that the speed v of all particles emerging at C must be $v = E/B$. Since this is independent of mass and charge, it follows that nothing can be concluded about the values for q and m for the particles that arrive at slit C. By varying the ratio E/B we can, in effect, control the velocities of the particles which arrive at C; thus the apparatus in Figure 26-4 is known as a *velocity selector* for charged particles. Neutral particles would go from A to C regardless of their velocities.

26-3 The mass spectrometer

As an application of the preceding formulas we now describe the principles underlying the instrument known as a *mass spectrometer*, which can be used to measure the mass of a charged particle such as a heavy ion. This device was invented and first constructed in 1919 by Francis W. Aston (1877–1945), who was interested in establishing the masses of the isotopes associated with the various chemical elements. Since its invention, many refinements of the mass spectrometer have been made, and today it continues to be one of the important tools of chemists and physicists.

Figure 26-5 shows the key elements of a mass spectrometer. A collimated beam of charged particles of various masses and velocities enters via slit system A a region of space in which there exist an electric and a magnetic

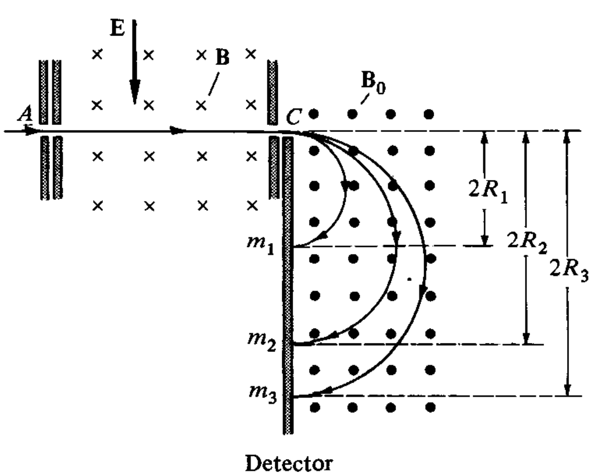


Figure 26-5

field, **E** and **B**, at right angles to each other. According to the results of Example 26-2, this configuration of fields acts as a *velocity selector*, and the magnitude of the common velocity **v** of the particles that get through to slit C is

$$v = \frac{E}{B}$$

Upon emerging from slit C, these particles of velocity **v** now enter a region where, as indicated by the dots, there is only a magnetic induction $\mathbf{B}_0$ perpendicular to and out of the plane of the diagram. The particles here traverse a semicircular path until they strike a screen or some other detecting instrument. According to (26-3), the distance $2R$ below the slit C at which a particle will be detected is

$$2R = \frac{2mv}{qB_0}$$

and substituting the known velocity $v = E/B$ this distance may be expressed in the equivalent form

$$2R = 2m\left(\frac{E}{qBB_0}\right) \tag{26-6}$$

Assuming that the charge q is known, since E, B, and B_0 are each separately measurable, we conclude that a measurement of the distance $2R$ enables us, in effect, to deduce the mass m of the particle. Alternatively, for purposes of measuring relative masses, as in the determination of isotopes, we may express (26-6) in the form

$$\frac{2R_1}{2R_2} = \frac{m_1}{m_2} \tag{26-7}$$

where $2R_1$ and $2R_2$ are the distances defined in Figure 26-5 and correspond, respectively, to the isotopes of masses m_1 and m_2. In words, (26-7) states that the distance below slit C that a particle strikes the detector in Figure 26-5 is directly proportional to its mass. The coefficient of proportionality, according to (26-6), is $(2E/qBB_0)$.

26-4 The cyclotron

As a second application of the formulas in Section 26-2 let us discuss briefly the principles underlying the operation of a cyclotron.

The cyclotron is an instrument that was invented and made operational in 1931 by E. O. Lawrence (1901–1958) and by means of which charged particles, such as protons or ionized helium atoms, can be accelerated to very high speeds. Although the cyclotron is a rather complex instrument, the basic principles underlying its operation are simple, and thus let us focus only on these.

The essential elements of a cyclotron are an electromagnet, which can

produce a **B**-field of the order of a few tesla, and two metal chambers, called "dees," which are shaped like the halves of a large pillbox cut in two. Figure 26-6a shows a side view of these elements, and the same arrangement is viewed from the top in Figure 26-6b. As shown, the two dees are connected to an oscillator so that the potential difference between them can be reversed at some fixed frequency of the order of 10^6 Hz.

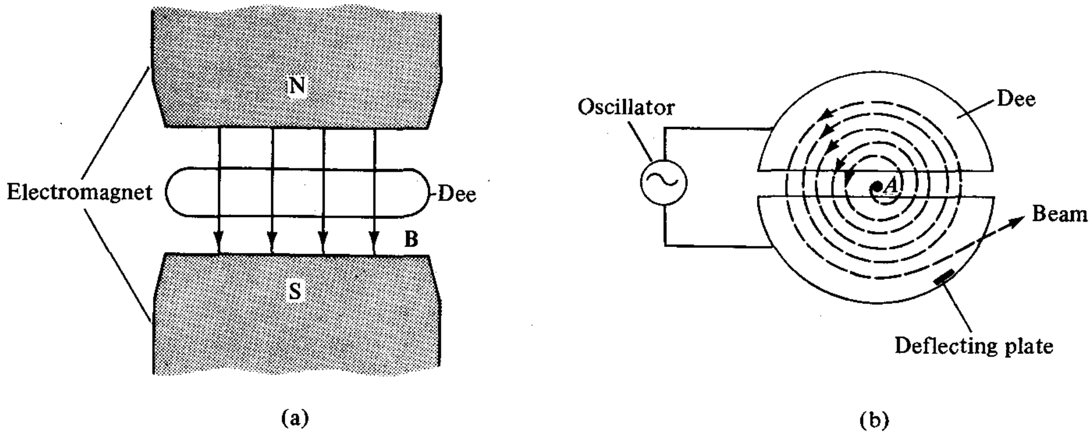


Figure 26-6

Suppose a positively charged ion is introduced at point A in Figure 26-6b. Assuming that it has a small velocity $\mathbf{v}_0$ perpendicular to **B**, the ion will traverse a semicircular orbit in the upper dee with a radius R_0 given, according to (26-3), by

$$R_0 = \frac{mv_0}{qB} \tag{26-8}$$

where m and q are, respectively, the mass and charge of the particle. Further, according to (26-5), the time τ required to traverse this path is independent of the velocity and has the constant value

$$\tau = \frac{1}{2}P = \frac{\pi m}{qB} \tag{26-9}$$

There is essentially no electric force acting on the particle while it is within the dee, since we know from studies in electrostatics that the electric field in the interior of a conducting shell vanishes.

After the particle has traversed this first semicircular orbit, it leaves the upper dee, and while in the gap between the dees it experiences an electric force due to the potential difference between the dees. As a result, assuming that the lower dee is at this instant at the lower potential, it follows that this ion will be accelerated and subsequently enter the lower dee with a velocity $\mathbf{v}_1$ larger than was its initial velocity $\mathbf{v}_0$. In the lower dee, because of the magnetic induction, it will again traverse a semicircular orbit but with a somewhat larger radius corresponding, in accordance with (26-3), to its new velocity $\mathbf{v}_1$. However, and this is a most important point, the time τ required

to traverse this orbit around the lower dee is the *same* as it was in the upper one; namely, the time τ in (26-9). If the frequency of the oscillator is set so that the potential difference between the dees is periodically reversed in this time interval τ, then each time the particle leaves one dee to go to the other it will accelerate. In other words, if the angular frequency ω_{os} of the oscillator is adjusted to be the same as that of the cyclotron frequency in (26-4)

$$\omega_{os} = \omega_c = \frac{qB}{m} \tag{26-10}$$

then each time the particle crosses between the dees it will be speeded up. The radius of its orbit will also increase each time because of the associated increases in velocity. Finally, when the orbital radius is sufficiently large, the particle is ejected from the cyclotron through an exit port by means of a charged deflecting plate.

To sum up, if the angular frequency of the oscillator connected to the dees is adjusted to be the cyclotron frequency, then the orbit of the particle consists of a series of semicircular paths of ever-increasing velocities and corresponding radii. The time τ required to traverse each semicircular orbit is given in (26-9) and is a constant independent of the velocity. Thus each time the ion traverses the gap its energy increases, until finally it is ejected by a deflecting plate at the exit port of the cyclotron.

To calculate the energy E of the particles accelerated in a cyclotron let R represent the "radius" of the cyclotron; that is, the radius of the largest orbit of an ion just prior to ejection. According to (26-3), the maximum velocity v_{max} of an accelerated ion is

$$v_{max} = \frac{qB}{m}R \tag{26-11}$$

Hence its energy, and that of the cyclotron, is

$$E = \frac{1}{2}mv_{max}^2 = \frac{1}{2m}q^2B^2R^2 \tag{26-12}$$

where the second equality follows by substitution from (26-11). Note that the energy of a cyclotron is directly proportional to B^2 and to R^2 and varies inversely with the mass m of the ion being accelerated.

Example 26-3 A cyclotron used to accelerate protons has a field of strength 0.8 tesla and a maximum orbital radius of 0.3 meter.

(a) What must be the angular frequency of the oscillator connected to the dees?
(b) What is the velocity of the protons after they are ejected?
(c) Calculate the final energy of the protons.
(d) What would the final energy be if deuterons were accelerated in this cyclotron?

Solution

(a) Making use of (26-10) we find that

$$\omega_{os} = \frac{qB}{m} = \frac{1.6 \times 10^{-19}\,\text{C} \times 0.8\,\text{T}}{1.67 \times 10^{-27}\,\text{kg}}$$
$$= 7.7 \times 10^7\,\text{rad/s}$$

where use has been made of the known values of q and m for a proton.

(b) On substituting the given values for q, B, m, and R into (26-11) we obtain

$$v_{\text{max}} = \frac{qBR}{m} = \frac{1.6 \times 10^{-19}\,\text{C} \times 0.8\,\text{T} \times 0.3\,\text{m}}{1.67 \times 10^{-27}\,\text{kg}}$$
$$= 2.3 \times 10^7\,\text{m/s}$$

(c) Substitution into (26-12) yields for the energy E

$$E = \frac{1}{2m}(qBR)^2$$
$$= \frac{1}{2}\frac{[1.6 \times 10^{-19}\,\text{C} \times 0.8\,\text{T} \times 0.3\,\text{m}]^2}{1.67 \times 10^{-27}\,\text{kg}}$$
$$= 4.4 \times 10^{-13}\,\text{J}$$
$$= 2.8\,\text{MeV}$$

where the final equality follows since by definition

$$1.6 \times 10^{-13}\,\text{J} = 1\,\text{MeV}$$

The final energy of 2.8 MeV means that the protons have the same energy as if they had been accelerated through a potential difference of 2.8 million volts.

(d) Since the mass of the deuteron is twice that of the proton, while their charges are the same, (26-12) shows that the final energy of an accelerated deuteron will be only half that acquired by a proton. Accordingly, making use of the result of (c) we find that

$$E = 1.4\,\text{MeV} \qquad \text{(deuterons)}$$

Example 26-4 A particular cyclotron produces 18-MeV alpha particles. If the frequency of the oscillator is 1.2×10^7 Hz, calculate:

(a) The strength of the magnetic induction.

(b) The number of times that an α particle orbits inside the cyclotron if its energy is increased by 0.15 MeV during each orbit.

Solution

(a) Solving (26-10) for B we obtain

$$B = \frac{m\omega_{os}}{q}$$

Now a frequency of 1.2×10^7 Hz corresponds to an angular frequency ω_{os}:

$$\omega_{os} = 2\pi \times 1.2 \times 10^7\,\text{rad/s}$$
$$= 7.5 \times 10^7\,\text{rad/s}$$

Hence

$$B = \frac{m\omega_{os}}{q} = \frac{6.67 \times 10^{-27}\ \text{kg} \times 7.5 \times 10^{7}\ \text{rad/s}}{3.2 \times 10^{-19}\ \text{C}}$$
$$= 1.6\ \text{T}$$

where we have used the fact that the charge of an α particle is twice that of a proton, while its mass is four times greater.

(b) Since the particle picks up 0.15 MeV during each full orbit and its final energy is 18 MeV, it follows that the number n of such traversals is

$$n = \frac{18\ \text{MeV}}{0.15\ \text{MeV}} = 120$$

Hence, since the frequency of the oscillator is 1.2×10^7 Hz, the total acceleration time is

$$\frac{120}{1.2 \times 10^{7}\ \text{Hz}} = 1.0 \times 10^{-5}\ \text{s}$$

and this is a very short time indeed.

Example 26-5 According to the theory of relativity, for speeds close to that of light ($c = 3 \times 10^8$ m/s), (26-2) must be replaced by

$$qvB = \frac{mv^2/R}{[1 - v^2/c^2]^{1/2}} \qquad \textbf{(26-13)}$$

(a) What is the cyclotron frequency in this case?

(b) How long does it take such a particle to traverse a dee?

(c) Explain why a cyclotron, as described above, does *not* work for relativistic particles.

Solution

(a) A comparison of (26-2) and (26-13) shows that if we make the replacement

$$m \to \frac{m}{[1 - v^2/c^2]^{1/2}}$$

in the former we obtain the latter. Hence, the relativistic generalization of (26-4) is

$$\omega_c = \frac{qB}{m}\left[1 - \frac{v^2}{c^2}\right]^{1/2}$$

and thus for high speeds the cyclotron frequency decreases as v increases.

(b) Making the same replacement in (26-9) we find that

$$\tau = \frac{\pi m}{qB}\frac{1}{[1 - v^2/c^2]^{1/2}}$$

(c) According to this formula for τ, the time a particle spends in a dee varies with its speed. Hence if an oscillator of angular frequency ω_{os} given in (26-10) is used, then, as the velocities of the accelerating particles approach the speed of light, they will not consistently cross the gap between the dees in phase with the oscillating field.

This means that they will not be consistently accelerated when crossing the gap; indeed, during some crossings they will be slowed down. Hence the effectiveness of a cyclotron is restricted to particle speeds small compared to the speed of light.

Because of this restriction, during recent decades much effort has gone into the design and construction of a variety of accelerators that are not subject to this limitation. Figure 26-7 shows one of these. This accelerator is 2 km in diameter and accelerates protons to a peak energy of about 300 GeV $= 3 \times 10^5$ MeV! That is, it can accelerate protons to energies about 1000 times greater than is possible for a cyclotron. As of this writing (1973) these are the most energetic particles produced, in a controlled way, by man.

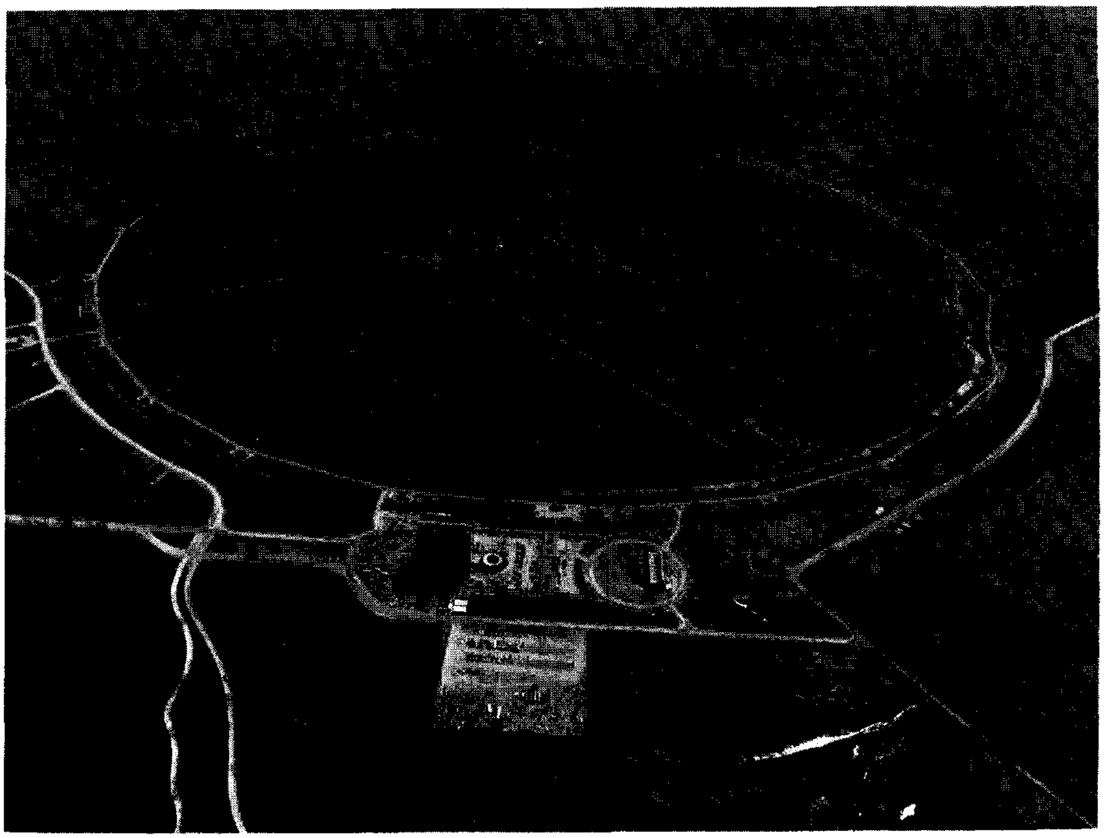


Figure 26-7 *Aerial view of the main accelerator at the Fermi National Accelerator Laboratory (FNAL) in Batavia, Illinois. The circumference of the accelerator is about 4 miles. (Courtesy FNAL.)*

26-5 The force on current

Consider, in Figure 26-8, an infinitesimal element of length dl of a wire of cross-sectional area A and through which flows a uniform current i. Suppose also the existence of an external **B**-field—that is, one produced by magnetic sources other than that due to this current i—and let $d\mathbf{F}$ represent the magnetic force on the current due to this **B**-field. According to the physical picture of macroscopic currents developed in Chapter 23, the current flow through the element in Figure 26-8 can be thought of as being due to a certain

number, n, of charge-carriers per unit volume, each of which travels at the drift velocity $\mathbf{v}_d$ along the direction of the current. The total charge ΔQ involved is therefore

$$\Delta Q = qnA\, dl$$

where q is the charge on each of these particles. Substituting this formula into (26-1), we may express $d\mathbf{F}$ as

$$d\mathbf{F} = qnA\, dl\, (\mathbf{e}v_d) \times \mathbf{B} \tag{26-14}$$

where $\mathbf{e}$ is a unit vector along the direction of the current. Moreover, according to (23-6) the current i is related to the above parameters by

$$i = qnAv_d$$

and therefore (26-14) may be reexpressed in the form

$$d\mathbf{F} = i\, dl\, \mathbf{e} \times \mathbf{B}$$

Finally, defining the vector $d\mathbf{l}$ to have the magnitude dl and to be directed along the current, that is, along the unit vector $\mathbf{e}$, we may write

$$d\mathbf{F} = i\, d\mathbf{l} \times \mathbf{B} \tag{26-15}$$

a formula which is basic for calculating the magnetic force on a wire that carries a current.

For the more general case of a closed current loop, it is necessary to add together contributions of the form in (26-15) for each current element $i\, d\mathbf{l}$ of the loop. Expressing this sum in the form of an integral, we find

$$\mathbf{F} = i \oint d\mathbf{l} \times \mathbf{B} \tag{26-16}$$

where $\mathbf{F}$ is the total force on the loop, and the circle superimposed on the integral means that the integral is to be carried out around the entire loop.

Example 26-6 A wire 2 meters long carries a current of 5.0 amperes and is in a uniform **B**-field of strength 0.03 tesla, which makes an angle of 30° with the wire. See Figure 26-9. Calculate the force on the wire.

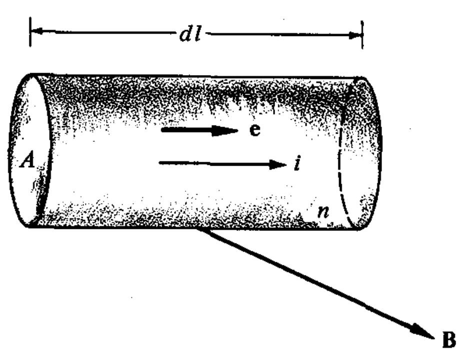


Figure 26-8

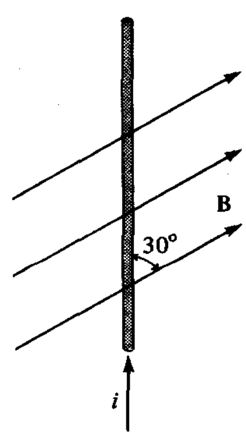


Figure 26-9

Solution According to the definition of a cross product and (26-15), the force on each element of the wire is directed perpendicular to and down into the plane of Figure 26-9. The magnitude dF of this force on an arbitrary element of length dl is

$$dF = |i\,d\mathbf{l} \times \mathbf{B}| = i\,dl\,B \sin 30°$$

Since i and B are constant, the magnitude of the *total* force on the wire is therefore

$$\begin{aligned} F &= ilB \sin 30° \\ &= (5.0\text{ A}) \times (2.0\text{ m}) \times (0.03\text{ T}) \times 0.5 \\ &= 0.15\text{ N} \end{aligned}$$

where in the second equality we have used $\sin 30° = 0.5$.

Example 26-7 Show that a closed loop of current in a uniform magnetic induction field experiences no force.

Solution Since the **B**-field is constant in this case, the factor **B** can be taken out from under the integral sign in (26-16), and it assumes the form

$$\mathbf{F} = i\left[\oint d\mathbf{l}\right] \times \mathbf{B}$$

To show that **F** vanishes, we shall now establish that the integral

$$\mathbf{I} = \oint d\mathbf{l}$$

vanishes over any closed path.

Consider, in Figure 26-10, the replacement of a closed loop by a sequence of N infinitesimal vectorial elements $d\mathbf{l}_i\,(i = 1, 2, \ldots, N)$. In the figure $N = 9$. Recalling that an integral is the limit of a sum, it follows that if the elements $\{d\mathbf{l}_i\}$ are sufficiently small, then

$$\mathbf{I} \cong \sum_{i=1}^{N} d\mathbf{l}_i \tag{26-17}$$

where the approximate equality becomes equality in the limit as each of the $\{d\mathbf{l}_i\}$ tends to zero. But the sum of a collection of vectors may be obtained graphically by placing the vectors in sequence with the head of one touching the tail of the next, with the sum then given by the vector connecting the head of the last to the tail of the first. For a closed loop, such as in Figure 26-10, the head of each vector touches the tail of another and in particular the head of the last touches the tail of the first. Hence their vector sum is zero. It follows that the sum in (26-17) vanishes, and thus passing to the limit of vanishingly small elements $\{d\mathbf{l}_i\}$ we conclude that the integral **I** in (26-17) does also. This establishes the fact that there is no magnetic force on a closed loop in a *uniform* magnetic induction field.

Example 26-8 A counterclockwise current i flows around the rectangular loop $ACDEA$ of dimensions $(b - a)$ and c, as shown in Figure 26-11. Calculate the force on this current loop due to an infinitely long wire, parallel to a side of the loop, and through which there flows a current I directed as shown.

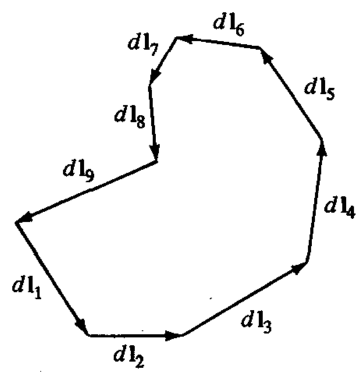


Figure 26-10

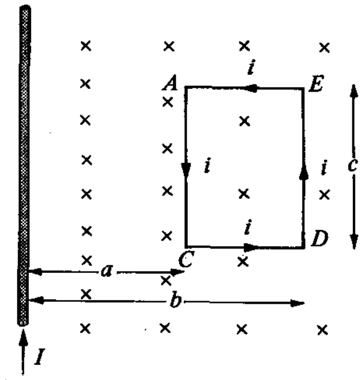


Figure 26-11

Solution The magnitude of the magnetic induction **B** produced by an infinitely long wire at a perpendicular distance r from the wire is, according to (25-6),

$$B = \frac{\mu_0 I}{2\pi r} \tag{26-18}$$

and as indicated by the crosses its direction is perpendicular and down into the plane of the figure.

Making use of the definition of a cross product and (26-15), we find that the force F_{AC} on the wire segment AC is directed toward the right and has the strength

$$F_{AC} = icB = \frac{ic\mu_0 I}{2\pi a}$$

where the second equality follows from (26-18). In a similar way the force F_{ED} on the segment ED has the magnitude

$$F_{ED} = \frac{ic\mu_0 I}{2\pi b}$$

but this is directed toward the left. Since the currents in the upper and lower segments AE and CD flow in opposite directions, the forces on these segments are equal and opposite; thus there is no contribution to the force due to them. Combining these results we obtain for the magnitude of the total force **F** on the entire loop

$$F = F_{AC} - F_{ED} = \frac{ic\mu_0 I}{2\pi}\left(\frac{1}{a} - \frac{1}{b}\right)$$

and this is directed to the right in the figure.

Note that since **B** is not uniform in this case, the force on the closed loop need not vanish as for the uniform field in Example 26-7.

26-6 Torque on a current loop in a uniform B-field

In Example 26-7 we saw that there is no net force on a current loop in a uniform **B**-field. Nevertheless, depending on the orientation of the current loop relative to **B**, there may be a torque on the loop, and the purpose of this section is to examine this possibility.

Consider, in Figure 26-12, a rectangular loop of sides a and b and carrying a current i. For convenience let us set up a coordinate system, with the loop lying in the y-z plane, and suppose first the existence of a uniform magnetic induction $\mathbf{B}$ along the x-axis, perpendicular to the plane of the loop. According to (26-15), the forces $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$, and $\mathbf{F}_4$ on each of the straight segments are directed as shown in the figure and have the magnitudes

$$|\mathbf{F}_1| = |\mathbf{F}_3| = ibB \qquad |\mathbf{F}_2| = |\mathbf{F}_4| = iaB \tag{26-19}$$

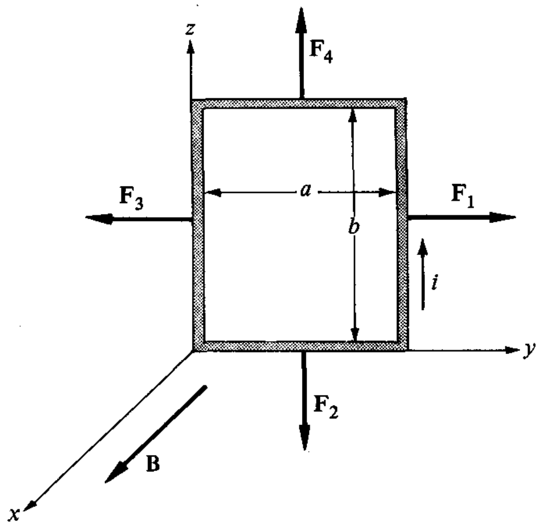


Figure 26-12

Consistent with the result of Example 26-7, these equations imply that there is no net force on the loop. Furthermore, since all four forces lie in the y-z plane and are pairwise equal and opposite, it follows that the net torque on the loop produced by these forces also vanishes.

Consider now the same physical situation as above, but with the $\mathbf{B}$-field no longer perpendicular to the plane of the loop. By analogy to the discussion in Section 25-9, let us define a unit vector $\mathbf{n}$ normal to the plane of the loop and with sense such that:

If the loop is grasped in the right hand with fingers pointing along the current, then the outstretched thumb points along $\mathbf{n}$.

With this convention, suppose in Figure 26-13 that $\mathbf{B}$ is directed along the x-axis, but now the normal $\mathbf{n}$ to the plane of the loop makes an angle α with the positive x-axis. Applying (26-15) to this system, we find that the forces $\mathbf{F}_1$ $\mathbf{F}_2$, $\mathbf{F}_3$, and $\mathbf{F}_4$ are directed as shown in the figure; that is, $\mathbf{F}_4$ and $\mathbf{F}_2$ are along the positive and negative z-axis, respectively, and correspondingly for $\mathbf{F}_1$ and $\mathbf{F}_3$ along the y-axis. Their magnitudes are precisely the same as in (26-19) and, again, there is no net force acting on the loop. However, this time, the forces $\mathbf{F}_1$ and $\mathbf{F}_3$ no longer lie in the plane of the loop. Hence $\mathbf{F}_1$ and $\mathbf{F}_3$ produce a

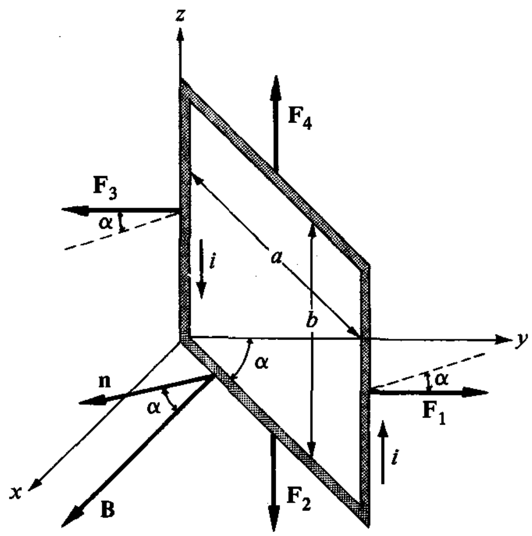


Figure 26-13

torque $\boldsymbol{\tau}$, about the center of the loop, which tends to line up the normal **n** to the plane of the loop with the direction of **B**. That is, the torque on the loop in Figure 26-13 is directed so that the loop tends to turn to the equilibrium arrangement in Figure 26-12.

Having established the direction of the torque on the loop, let us calculate its magnitude. To this end, let us view the loop in Figure 26-13, but this time along the direction of the positive z-axis, as in Figure 26-14. According to (26-19), the forces $\mathbf{F}_1$ and $\mathbf{F}_3$ both have the magnitude ibB, and thus they contribute equally to the torque about the axis through the center of the loop and parallel to the z-axis. Because of the fact that the torque $\boldsymbol{\tau}$ produced by a force **F** is

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$$

where **r** is a vector from the origin to the point where the force acts, it follows that the magnitude of the torque produced by each of $\mathbf{F}_1$ and $\mathbf{F}_3$ is

$$|\mathbf{r} \times \mathbf{F}_3| = |\mathbf{r} \times \mathbf{F}_1| = \frac{a}{2}|\mathbf{F}_1| \sin\alpha = \frac{iab \sin\alpha}{2} B$$

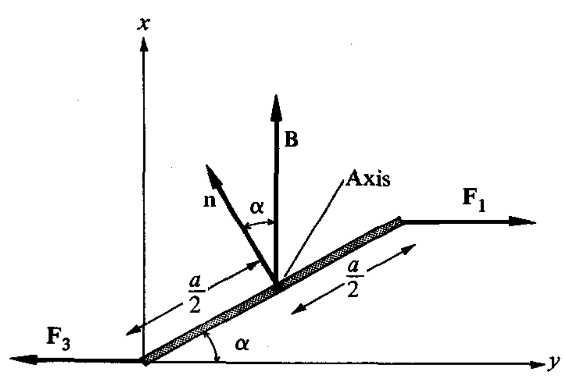


Figure 26-14

Hence the magnitude of the total torque τ on the loop in Figure 26-13 is

$$\tau = iabB \sin \alpha \tag{26-20}$$

and its direction is along the positive z-axis. That is, the direction of this torque acts in such a way that the normal $\mathbf{n}$ to the plane of the loop tends to line up with the direction of the magnetic induction.

Although (26-20) has been derived only for the special case of a rectangular loop, in a certain modified form, as established in the problems, it is true in general. Since the area A of a rectangular loop is ab, (26-20) may be expressed in the equivalent form

$$\tau = iAB \sin \alpha \tag{26-21}$$

a formula which is valid also for loops of arbitrary shape. As above, this torque tends to line up the normal $\mathbf{n}$ to the plane of the loop along $\mathbf{B}$.

If there are N turns of wire in the loop, then effectively it carries a current Ni. The generalization of (26-21) to this more general case is

$$\tau = iNAB \sin \alpha \tag{26-22}$$

To obtain a vector form for (26-22), consider, in Figure 26-15, a planar loop of N turns, area A, and current i, and let $\mathbf{n}$ be a unit vector normal to the loop with its sense determined as above. If we define the *magnetic dipole moment* $\boldsymbol{\mu}$ of this loop to be a vector directed along the normal $\mathbf{n}$ and to have the magnitude iNA, then

$$\boldsymbol{\mu} = iNA\mathbf{n} \tag{26-23}$$

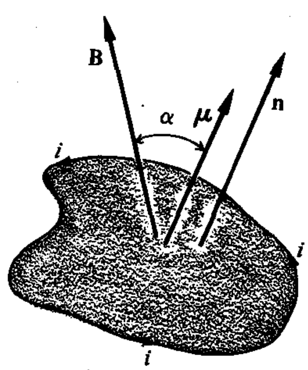


Figure 26-15

and the torque $\boldsymbol{\tau}$ acting on the current loop may be expressed as

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B} \tag{26-24}$$

It is easily verified that this formula gives correctly both the magnitude and the direction of $\boldsymbol{\tau}$; it is analogous to the formula for the torque acting on a dipole $\mathbf{p}$ in an electric field $\mathbf{E}$; that is,

$$\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E}$$

as derived in Problem 6, Chapter 20.

Example 26-9 A thin, circular loop of wire of radius 10 cm carries a current of 5.0 amperes. Calculate:

(a) The magnetic dipole moment of the loop.

(b) The torque on it if it is in a uniform field of strength 0.2 tesla and making an angle of $\pi/4$ with the normal to the loop.

Solution

(a) Since the area A of the loop is

$$A = \pi r^2 = \pi(0.1\ \text{m})^2 = 3.14 \times 10^{-2}\ \text{m}^2$$

we find, by use of (26-23), that

$$\begin{aligned}\boldsymbol{\mu} &= \mathbf{n}iA \\ &= \mathbf{n}(5.0\ \text{A}) \times (3.14 \times 10^{-2}\ \text{m}^2) \\ &= \mathbf{n}0.16\ \text{A-m}^2\end{aligned}$$

where $\mathbf{n}$ is the unit vector normal to the plane of the loop with its sense defined above.

(b) The magnitude of $\boldsymbol{\tau}$ may be obtained, by use of (26-24) and (26-23), as

$$\begin{aligned}\tau &= |\boldsymbol{\mu} \times \mathbf{B}| = \mu B \sin\alpha \\ &= (0.16\ \text{A-m}^2) \times (0.2\ \text{T}) \times 0.71 \\ &= 2.3 \times 10^{-2}\ \text{N-m}\end{aligned}$$

since $\sin(\pi/4) = 0.71$.

26-7 The galvanometer

In Chapter 24 we introduced the *galvanometer* as a device that can be used to measure current. The purpose of this section is to discuss briefly the principles underlying its operation.

First, however, let us consider an early form of the galvanometer developed by Oersted. Suppose, in Figure 26-16, a compass is placed in a horizontal position at a distance d above a very long wire oriented along the

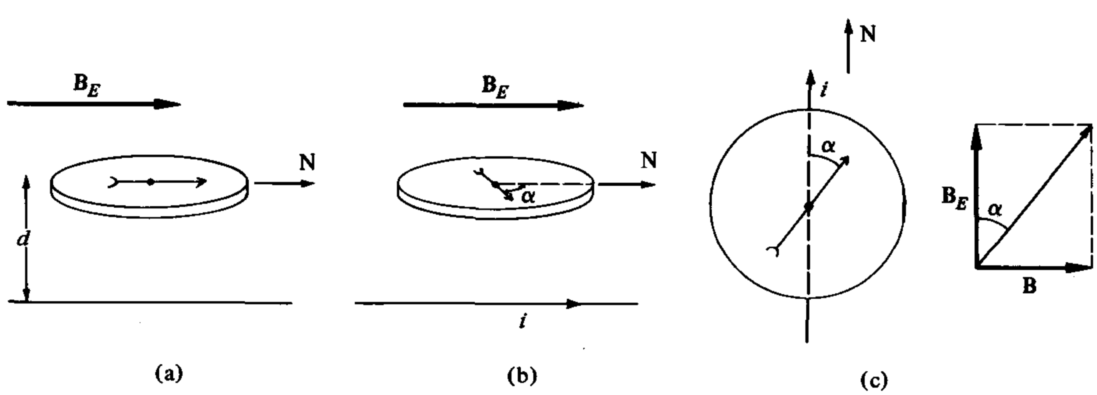


Figure 26-16

north-south direction. In the absence of any current in the wire, the compass needle will line up with the horizontal component $\mathbf{B}_E$ of the earth's magnetic induction. As shown, in this circumstance it will point north.

Suppose now a current i flows north along the wire. Then in addition to the northerly earth-field $\mathbf{B}_E$, there will also be a magnetic induction $\mathbf{B}$ directed eastward and of strength

$$B = \frac{\mu_0 i}{2\pi d}$$

The compass needle will point along the direction of the total field and thus, as shown in Figure 26-16b, it will now point in a direction which makes a certain angle α with respect to north. Reference to the top view of this situation in Figure 26-16c shows that

$$\tan\alpha = \frac{B}{B_E} = \frac{\mu_0 i}{2\pi d B_E}$$

and solving for i yields

$$i = \frac{2\pi d B_E}{\mu_0}\tan\alpha$$

Finally, since all quantities on the right-hand side are known or independently measurable, this instrument effectively measures current.

There are several practical difficulties associated with the above galvanometer and it is not used in practice. Most galvanometers in use today are of the *d'Arsonval* moving-coil or pivoted-coil type. These galvanometers contain a movable coil of wire, in a fixed **B**-field, and the current to be measured is allowed to pass through the coil. A measurement of the associated torque on the coil thus enables us, by use of (26-22), to measure the desired current.

Figure 26-17 shows the main elements of a galvanometer of the pivoted-coil type. A flat, rectangular coil of wire, which will carry the current to be measured, encircles a fixed, soft-iron core, and is suspended in a way so that it is free to rotate about an axis perpendicular to the plane of the figure. The coil

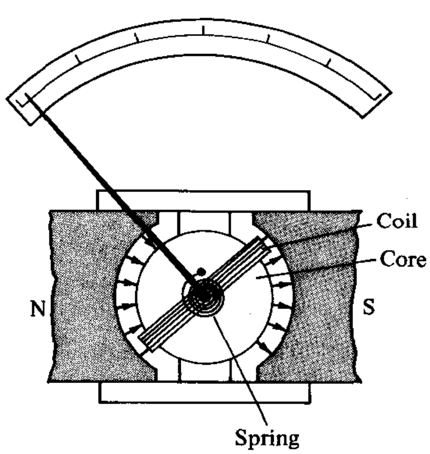


Figure 26-17

itself is maintained in a fixed equilibrium position by a spring, and a pointer is attached so that rotations about the axis are easily observed. As shown, the core is fixed between the cylindrically concave pole tips of a permanent magnet and helps to produce the required radial **B**-field. If a current of strength i flows through the galvanometer, then the coil will experience a torque, which, according to (26-22), is proportional to i. For the direction of current shown, the coil and its attached pointer will move clockwise until this motion is counteracted by the torque produced by the stretched spring. Assuming the latter torque to be proportional to the angle of rotation of the coil, it follows that this angle of rotation is also proportional to the current. In other words, the angle through which the pointer moves is directly proportional to the current, so that once the galvanometer has been calibrated at some known value of the current it may be used to measure the strength of other currents.

The roles played by the soft-iron core and the shapes of the poles of the permanent magnet should be carefully noted. As will be clarified in Chapter 28, the iron-core concentrates the field lines and thereby helps to produce a nearly radial and uniform **B**-field at the position of the coil windings. The angle α between **B** and the normal to the plane of the coil is then always 90° and the factor $\sin \alpha$ $(= 1)$ in (26-22) is the same for all currents. Hence the torque on the coil, (and the associated angular displacement of the pointer) is proportional to the current, with the same coefficient of proportionality for all angles.

26-8 The Hall effect

A striking illustration of the fact that a magnetic induction field exerts a force on current is provided by a phenomenon discovered by E. H. Hall in 1879 and now known as the *Hall effect*. This effect is very important in the experimental studies of the electrical properties of matter and can be readily understood at an elementary level.

Consider, in Figure 26-18a, a thin strip of material of cross-sectional area A, height h, and length l. If a voltage source (not shown in the figure) is connected between the end faces of the strip, an electric field **E** directed as shown will appear inside and associated with it will be a certain current i. If further, a voltmeter is connected transversely across the strip, it will give a zero reading since the electric field has no component along the vertical direction. The electrons in the sample will simply move with their characteristic drift velocity $\mathbf{v}_d$ in a direction opposite to the field.

Suppose now, as shown in Figure 26-18b, the existence of a magnetic induction field **B** directed perpendicular to the slab. In response to this field, the electrons will no longer move along the direction of $-\mathbf{E}$, but instead they will move downward in a curved path. As a consequence of this motion, a piling up of negative charge takes place along the bottom of the strip and a compensating positive charge appears at the top. As shown in Figure 26-18c,

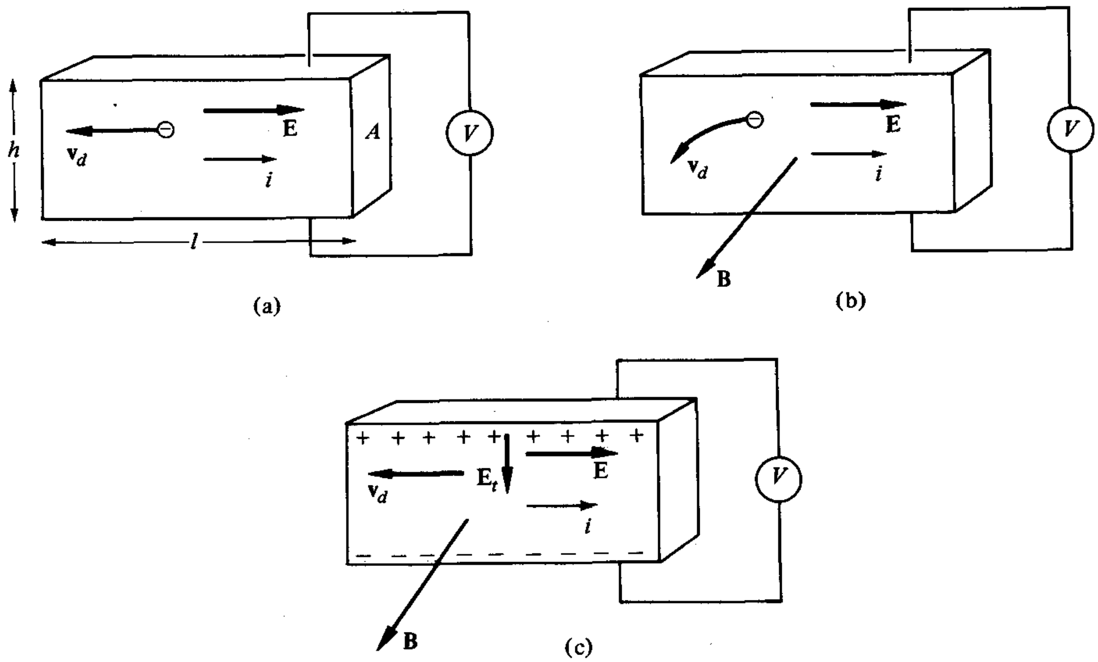


Figure 26-18

this piling up of charge will continue until its associated transverse electric field $\mathbf{E}_t$ cancels out the effect of the magnetic force. In other words, if q is the charge on an electron, then this transverse electric field $\mathbf{E}_t$ must satisfy

$$q\mathbf{E}_t + q\mathbf{v}_d \times \mathbf{B} = 0$$

for only then will this accumulation of charge cease. Taking into account the directions of $\mathbf{E}_t$, $\mathbf{v}_d$, and $\mathbf{B}$ in the figure, we find on canceling the common factor of q that

$$E_t = v_d B \tag{26-25}$$

The existence of this transverse field $\mathbf{E}_t$ is readily confirmed by virtue of the fact that for $B \neq 0$ the voltmeter in the figure will give a nonzero reading. The value of E_t itself is given by

$$E_t = \frac{V}{h}$$

where V is the reading on the voltmeter and h is the height of the slab. Since B can be measured independently, it follows from (26-25) that the Hall effect can be used to measure the drift velocity of the charge carriers in any material.

Additional information on the properties of the charge carriers can also be ascertained by measuring the strength of the current i flowing in the slab. Let us recall in this connection the relation (23-6)

$$i = nqAv_d$$

which relates the current i in a conductor to the number of charge carriers

per unit volume n, to the cross-sectional area A, and to the drift velocity. Solving this relation for v_d and substituting into (26-25), we obtain

$$E_t = \left(\frac{1}{nq}\right)\frac{iB}{A} \tag{26-26}$$

The coefficient $1/nq$ in this formula is called the *Hall coefficient* for the given material. Since i, E_t, B, and A are independently measurable, it follows that the Hall coefficient for a given substance can also be readily ascertained.

By use of (26-26), Hall coefficients for many metals have been measured. The fact that the charge carriers have a negative sign, and are thus presumably electrons, can also be determined, as in Figure 26-18c, by observing the direction of the transverse electric field $\mathbf{E}_t$. If, as shown there, the charge carriers are electrons, then a negative charge accumulates at the bottom of the slab, and $\mathbf{E}_t$ points downward. A most interesting fact that was discovered in this connection was that for some materials the direction of $\mathbf{E}_t$ is opposite to that in the figure. This implies that the charge carriers are positive for these cases! This remained something of a paradox until it was ultimately explained by the quantum theory of matter.

Example 26-10 A strip of copper is placed into a magnetic induction of strength 5.0 T (see Figure 26-18c), and a measurement of E_t yields the value 3.0×10^{-3} V/m.

(a) Calculate v_d.

(b) Assuming that $n \cong 8.0 \times 10^{28}$ electrons per cubic meter for copper and that the cross-sectional area $A = 5.0 \times 10^{-6}\ \text{m}^2$, calculate the current that flows along the strip.

(c) What is the Hall coefficient, $1/nq$?

Solution

(a) Substitution into (26-25) yields

$$v_d = \frac{E_t}{B} = \frac{3.0 \times 10^{-3}\ \text{V/m}}{5.0\ \text{T}}$$
$$= 6.0 \times 10^{-4}\ \text{m/s}$$

(b) Solving (26-26) for i,

$$i = nqA\frac{E_t}{B} = nqAv_d$$

and substituting the given data we find that

$$i = (8.0 \times 10^{28}/\text{m}^3) \times (1.6 \times 10^{-19}\ \text{C}) \times (5.0 \times 10^{-6}\ \text{m}^2) \times (6.0 \times 10^{-4}\ \text{m/s})$$
$$= 38\ \text{A}$$

(c) Substituting the given values for q and n we obtain

$$\frac{1}{nq} = \frac{1}{(8.0 \times 10^{28}/\text{m}^3) \times (-1.6 \times 10^{-19}\ \text{C})} = -7.8 \times 10^{-11}\ \text{m}^3/\text{C}$$

26-9 The ampere and the coulomb

In our discussion of Coulomb's law, the unit of charge, the coulomb, was defined as the amount of charge transported in 1 second through a wire in which there flows a current of 1.0 ampere. In order to complete this definition of the coulomb we now give an operational definition of the ampere.

To this end consider, in Figure 26-19, two very long and thin parallel wires, A and C, through which flow the respective currents i_A and i_C. If d is the separation distance between them, the magnetic induction $\mathbf{B}_A$ at wire C due to the current i_A has the magnitude

$$B_A = \frac{\mu_0 i_A}{2\pi d}$$

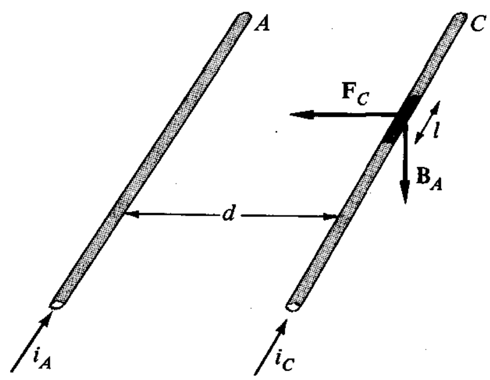


Figure 26-19

and, in accordance with the right-hand rule, its direction is as shown in the figure. Substituting this value for B_A into (26-15), we find that the force $\mathbf{F}_C$ on a length l of the wire has the magnitude

$$F_C = i_C l B_A = \frac{\mu_0 i_A i_C}{2\pi d} l \qquad \textbf{(26-27)}$$

and its direction is perpendicular to wire C and directed toward wire A. Correspondingly, the force on wire A due to i_C is equal and opposite to this force, and thus has the same magnitude as in (26-27).

In terms of (26-27) we now define the ampere as follows:

An ampere is the amount of current that, if flowing in each of two very long, parallel wires separated by a distance of 1 *meter, will cause each wire to experience a force of precisely* 2.0×10^{-7} N *for each meter of wire.*

Using the values $i_A = i_C = 1$ ampere, $d = 1$ meter, and $l = 1$ meter, we find by substitution into (26-27) that

$$F_C = \frac{\mu_0 i_A i_C l}{2\pi d} = \frac{(4\pi \times 10^{-7}\ \text{Wb/A-m}) \times (1.0\ \text{A}) \times (1.0\ \text{A}) \times (1.0\ \text{m})}{2\pi \times 1.0\ \text{m}}$$
$$= 2.0 \times 10^{-7}\ \text{N}$$

thus confirming the consistency of this definition with the preceding formulas.

In practice, it is not convenient to measure the ampere by use of two very long, straight wires, as above, but rather to make use of two parallel coils of wire separated by a few centimeters. The force of attraction, or repulsion—depending upon whether the currents are parallel or antiparallel—is then measured by a spring balance. At the National Bureau of Standards, precise measurements of currents are carried out in this way by a device known as a *current balance.*

To sum up, then, the ampere can be defined in terms only of measurements of force and of length by means of an apparatus housed at the Bureau and known as a current balance. Once the value of the ampere has been standardized in this way, the coulomb is defined as the flow of charge in 1 second through a wire carrying a current of 1 ampere.

26-10 Summary of important formulas

A particle of charge q and mass m traveling at a velocity $\mathbf{v}$ perpendicular to a uniform **B**-field will orbit about the field with a constant speed v in a circle of radius R:

$$R = \frac{mv}{qB} \qquad (26\text{-}3)$$

The plane of the orbit is perpendicular to **B**.

The force $d\mathbf{F}$ on a current element $i\,d\mathbf{l}$ of length dl in a magnetic induction **B** is

$$d\mathbf{F} = i\,d\mathbf{l} \times \mathbf{B} \qquad (26\text{-}15)$$

The torque $\boldsymbol{\tau}$ on a loop of N turns, area A, and carrying a current i is

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B} \qquad (26\text{-}24)$$

where **B** is the external field and where $\boldsymbol{\mu}$ is its magnetic dipole moment

$$\boldsymbol{\mu} = \mathbf{n}NiA \qquad (26\text{-}23)$$

with **n** a unit vector defined in Figure 26-15.

QUESTIONS

1. Define or describe briefly what is meant by the following: **(a)** cyclotron; **(b)** mass spectrometer; **(c)** MeV; and **(d)** Hall effect.
2. A proton in the upper reaches of the atmosphere is traveling in a direction parallel to the lines of the earth's magnetic induction. Will it experience a force? If it were traveling due east at some instant, what would be the direction of the force on it?
3. Suppose that you look down on a cyclotron and find that the lines of magnetic induction are directed toward you. Would you observe the protons to be traveling clockwise or counterclockwise?
4. Could a cyclotron be used to accelerate electrons? Why is this *not* done in practice? (*Hint*: See (26-13).)
5. Before entering a mass spectrometer, the particles are usually accelerated by a few kilovolts after they emerge from the ion source. Why?
6. A charged particle is at some instant traveling parallel to a long, straight wire in which there flows a current. Does it experience a force?
7. A charged particle is traveling at right angles to a straight wire carrying a current. Does it experience a force? Explain.
8. How could you prove experimentally that the charged particles in a cathode-ray-tube beam are negatively charged?
9. Describe a procedure by means of which it is possible to determine the sign of the charge carriers in any substance in which a current can be caused to flow.
10. Can you think of a way that it might be possible physically for the charge carriers in certain conductors to have a positive charge?
11. It is desired to change the direction of motion of a proton beam by an angle of 60°. Assuming that the protons are *monoenergetic*, that is, they all have the same velocity, describe an experimental setup by means of which this beam deflection can be accomplished.
12. A straight wire carrying a current i lies on the axis of a solenoid which carries a current of the same strength. Is there a force on the solenoid? Is there a torque?
13. A long solenoid that carries a current i is placed into a uniform magnetic induction. Does it experience a force? Will the solenoid exert a torque on the currents that produce the original uniform field?
14. According to (26-22) and (26-24), the torque on a current loop exerted by a uniform induction will vanish if the normal to the loop $\mathbf{n}$ is either parallel or antiparallel to $\mathbf{B}$. Describe in physical terms why it is that nevertheless the loop tends to line up preferentially with $\mathbf{n}$ parallel to $\mathbf{B}$.
15. To what extent do you think (26-26) is applicable if the magnetic induction is not uniform?
16. Need the force on a closed current loop in a *nonuniform* magnetic induction vanish? Explain.
17. Why is it *not* possible to design a galvanometer as in Figure 26-17 for which the pointer can swing through 360°? (*Hint*: Consider the **B**-field lines in the figure and the implications of this if the pointer had this degree of freedom.)
18. What do you suppose is the justification for our calling the quantity μ in (26-23) the magnetic dipole moment of a current loop?
19. In our studies of electrostatics we saw that it is not possible for an electric field to penetrate into the

interior of a charge-free conductor. Does the same feature hold for the magnetic induction? Would a cyclotron be able to operate if the **B** lines could not penetrate conducting shells?

20. In the discussion of magnetic forces on wires we actually calculated the force on the *current* in the wire and *not* on the wire itself. Explain why it is that our formulas for the magnetic force can be applied directly to the wire itself.

21. Suppose that in the apparatus in Figure 26-18 a slab of circular cross section instead of the rectangular one were used. Would there still be a Hall effect? Explain why a slab of rectangular cross section is to be preferred.

PROBLEMS

1. A proton in the upper reaches of the atmosphere is traveling at a velocity of 10^6 m/s at right angles to the earth's field, which has a strength of 5.0×10^{-5} tesla at that point.
   (a) What is the radius of the proton's orbit?
   (b) How long does it take the proton to complete one orbit?
   (c) What is its cyclotron frequency?
2. An electron is traveling at right angles to a uniform magnetic induction of strength 3.0×10^{-2} tesla. If its energy is 50 keV (1 keV = 10^{-3} MeV), calculate (a) its velocity; (b) the radius of its orbit; and (c) the cyclotron frequency.
3. A proton and an alpha particle are traveling in parallel directions at the same speed when they enter a region of space where there is a uniform magnetic induction **B**. Assume that they travel at right angles to the field.
   (a) What is the ratio of the radii of their orbits?
   (b) What is the ratio of their cyclotron frequencies?
   (c) What is the ratio of their energies?
4. A 2-MeV proton is traveling in a region of space where there is a uniform electric field of strength 10^5 V/m and a uniform magnetic induction **B** at right angles to it. If the direction of motion of the proton is perpendicular to the direction of both the electric field and the magnetic induction, and the proton is *not* accelerated, calculate the strength and the sense of direction of the **B**-field.
5. A beam containing a mixture of the isotopes ^{6}Li and ^{7}Li enters the region of the uniform magnetic induction B_0 via slit C in the magnetic spectrometer in Figure 26-5. If the ^{6}Li ions are detected at a distance of 10 cm below slit C, where will the ^{7}Li ions appear? What is the ratio of the kinetic energies of these two isotopes?
6. A cyclotron that is used to accelerate protons has a radius of 0.5 meter and a magnetic introduction of 0.75 tesla.
   (a) What is the energy of the emerging protons? Express your answer in joules and in MeV.
   (b) What is the final velocity of the ejected protons?
   (c) What would be the energies of α particles if they were accelerated by this cyclotron?
7. Consider the cyclotron of Problem 6.
   (a) What is the oscillator frequency

ω_{os} for this cyclotron when it accelerates protons?

(b) If the protons pick up 100 keV each time they cross the space between the dees, how many semicircular orbits will the protons complete before they are ejected?

(c) Calculate the time required to accelerate the protons up to their final velocities.

8. A cyclotron with a magnetic induction of 2.0 teslas is used to accelerate protons. (a) What must the frequency (in Hz) of the oscillating field between the dees be? (b) If this cyclotron is used to accelerate deuterons, to what must the frequency of this oscillating field be set?

9. If the cyclotron of Problem 6 were used to accelerate electrons, then, neglecting relativistic effects, what would be the final energies of the electrons? Are we justified in assuming that relativistic effects are negligible?

10. If E_0 is the energy of a cyclotron when it accelerates protons, show that it can accelerate ions, of atomic mass A and with Z units of charge, to the energy

$$E = \left(\frac{Z^2}{A}\right) E_0$$

11. If a magnetic monopole of strength ϵ existed, then the magnetic induction associated with it would be

$$\mathbf{B} = \left(\frac{\epsilon}{r^3}\right) \mathbf{r}$$

where $\mathbf{r}$ is the position in space measured from the monopole.

(a) Write down the equation of motion for a particle of charge q and mass m moving in the field of a monopole.

(b) Show that the kinetic energy of the particle is a constant of the motion.

*12. For the physical system in Problem 11 show that:

(a) The angular momentum $m\mathbf{r} \times \mathbf{v}$ (relative to the monopole) of the particle is *not* constant in general.

(b) The quantity $d(m\mathbf{r} \cdot \mathbf{v})/dt$ is constant, and determine its value in terms of the initial velocity $\mathbf{v}_0$.

13. It has been estimated that at the surface of a *neutron star* the **B**-field may be as high as 10^9 teslas. For a 10-MeV proton moving in such a field find (a) its cyclotron frequency and (b) the radius of its orbit.

14. A rectangular coil of wire carries a current i and has a width a. If the lower end of the coil is held between the poles of an electromagnet of strength B and directed as shown in Figure 26-20, calculate the magnitude of the force **F**, beyond that of gravity, required to support the coil. (*Note*: By suspending a coil from the arm of an analytical balance, the principle illustrated here has been used at the National Bureau of Standards for making precise measurements of magnetic fields.)

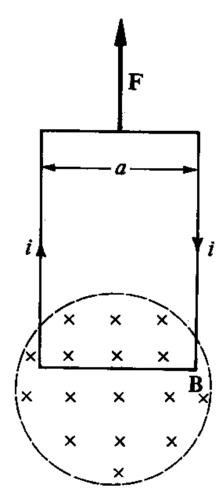


Figure 26-20

15. By use of the method of Example 26-7 show that the force on a portion of a wire carrying a current in a uniform magnetic induction is the same for all wires having the same endpoints. That is, show that the force on the wires 1 and 2 in Figure 26-21 are the same if the same currents flow in each.

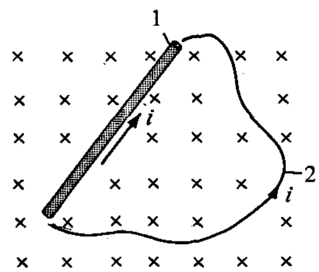


Figure 26-21

16. A current of 5.0 amperes flows in a square loop of side 10 cm. Calculate the total force on two of its adjacent sides produced by an external magnetic induction perpendicular to the plane of the loop and of strength 0.1 tesla.

17. A current of strength i flows in a coil in the shape of a regular pentagon of side a. Assume that there is a uniform induction of strength B perpendicular to the plane of the coil. **(a)** Calculate the force produced by the external field on each segment. **(b)** By making use of the results of (a), show explicitly that the total force on the loop is zero.

***18.** Consider two parallel wires each of length l, carrying the same currents i and separated by a distance a, as in Figure 26-22.

(a) Show that the magnetic induction **B** at the point (x, a) of the upper wire due to the current in the lower one has the magnitude

$$B = \frac{\mu_0 i}{4\pi}\frac{1}{a}\left\{\frac{(l/2)-x}{[(x-l/2)^2+a^2]^{1/2}} + \frac{(l/2)+x}{[(x+l/2)^2+a^2]^{1/2}}\right\}$$

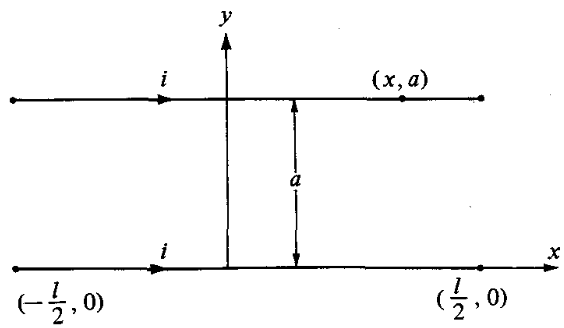


Figure 26-22

What is the direction of **B**?

(b) Calculate the force $d\mathbf{F}$ on the element of length dx located at the point (x, a) of the upper wire.

(c) Show by integration that the total force **F** on the upper wire is downward and has the strength

$$F = \frac{\mu_0 i^2}{2\pi}\left(\frac{1}{a}\right)[(l^2+a^2)^{1/2}-a]$$

19. Show that for appropriate values of the various parameters, the result in (c) of Problem 18 is consistent with (26-27).

20. In Figure 26-11, if $a = 3.0$ cm, $b = 5.0$ cm, $c = 3.0$ cm, $i = 5.0$ amperes, and $I = 2.0$ amperes, calculate **(a)** the force on segment AC due to the current in the long wire and **(b)** the force on segment AE due to the current in the long wire.

21. A rectangular loop of wire of sides 10 cm and 30 cm carries a current of 15 amperes. Make use of the result of Problem 18 to calculate the mutual forces of repulsion between the two pairs of opposite wires.

***22.** Consider the two wires of length l and L at right angles to each other in Figure 26-23. Assume that currents i and I are directed as shown.

(a) Show that the magnetic induction **B** at the position of the element of length dy located at

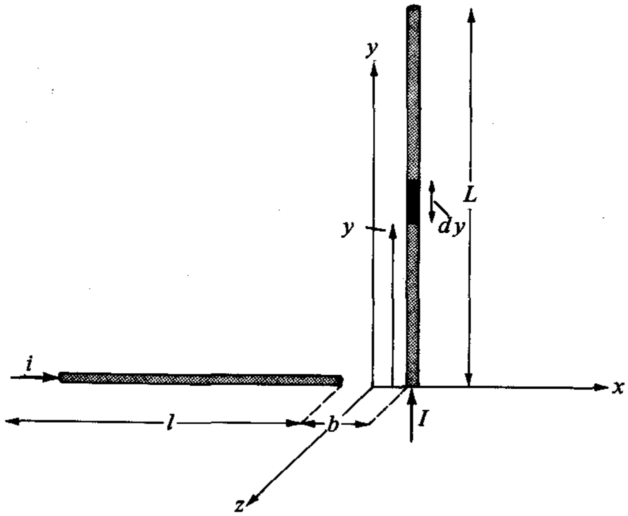


Figure 26-23

the point (0, y, 0) is

$$\mathbf{B} = \mathbf{k}\frac{\mu_0 i}{4\pi y}\left\{\frac{-b}{(b^2+y^2)^{1/2}} + \frac{b+l}{[(b+l)^2+y^2]^{1/2}}\right\}$$

where **k** is a unit vector along the z-axis.

(b) Show that the magnetic force dF on the element dy is

$$d\mathbf{F} = \mathbf{i}\frac{\mu_0 iI}{4\pi y}\,dy\left\{\frac{-b}{(b^2+y^2)^{1/2}} + \frac{b+l}{[(b+l)^2+y^2]^{1/2}}\right\}$$

where **i** is a unit vector along the x-axis.

(c) Calculate the total force **F** on the wire of length L.

23. Consider again the system in Figure 26-11. This time calculate the force on the long, straight wire due to the magnetic induction produced by the rectangular loop. Should your answer be the same as that given in Example 26-8?

24. Two current elements $i_1\,d\mathbf{l}_1$ and $i_2\,d\mathbf{l}_2$ are positioned relative to each other as shown in Figure 26-24.

(a) Calculate the force that the element $i_1\,d\mathbf{l}_1$ exerts on the other.

(b) Calculate the force which the element $i_2\,d\mathbf{l}_2$ exerts on the other element.

(c) Explain why, even though your results to (a) and (b) are not equal and opposite, there is no essential contradiction with Newton's law of action and reaction.

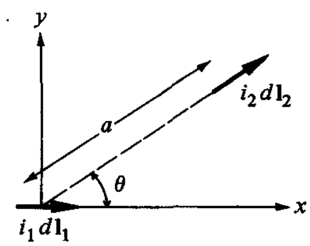


Figure 26-24

***25.** In a certain coordinate system a current I flows along an infinitely long wire that lies along the x-axis. Show that the magnetic force on a second wire of length l and carrying a current i whose endpoints are at the points $(0, 0, a)$ and $(0, l, a)$ is

$$\mathbf{F} = -\mathbf{i}\frac{\mu_0 iI}{4\pi}\ln\left[1+\frac{l^2}{a^2}\right]$$

26. Consider the same situation as in Problem 25, but this time suppose

the endpoints of the shorter segment are at the points (0, $-l/2$, a) and (0, $l/2$, a). (a) Show that this time there is no net force on the shorter segment. (b) Calculate the torque about the point $(0, 0, a)$ on the shorter segment.

27. Calculate the magnetic dipole moments associated with each of the following planar loops (assume that in each case the current is 2.0 amperes and that there are 10 turns in each loop):
(a) A circular loop of radius 10 cm.
(b) A rectangular loop of sides 2 cm and 10 cm.
(c) An elliptical-shaped loop of semimajor axis 10 cm and semiminor axis 5 cm.

28. By calculating the work required to rotate a current loop of dipole moment $\boldsymbol{\mu}$ in a magnetic induction **B**, show that the energy U associated with it is

$$U = -\boldsymbol{\mu} \cdot \mathbf{B}$$

(*Hint*: The work dW required to rotate a dipole by a small angle $d\alpha$ is $\tau\, d\alpha$, where τ is the torque that must be applied.)

29. A circular coil of wire of radius 10 cm and 150 turns carries a current of 10^{-2} ampere. What is the maximum torque that can be exerted on this coil by a uniform magnetic induction of strength 0.2 tesla?

30. A rectangular loop of sides a and b is suspended so that it is free to rotate about the horizontal axis AB (see Figure 26-25). If it has a mass

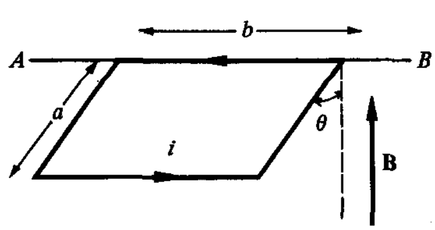


Figure 26-25

m and if the current around it is i, calculate the angle θ at which it will be in equilibrium in the presence of a uniform vertical magnetic induction **B**.

***31.** Consider, in Figure 26-26, a planar loop of wire carrying a current i in the presence of a uniform magnetic induction **B**.

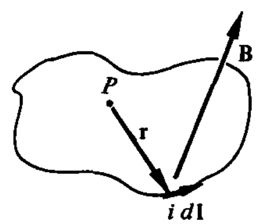


Figure 26-26

(a) Show that the torque $\boldsymbol{\tau}$ about a point P inside the loop may be expressed, by use of (26-15), in the form

$$\boldsymbol{\tau} = i \int \mathbf{r} \times (d\mathbf{l} \times \mathbf{B})$$

where **r** is the vector from the point P to the element $i\,d\mathbf{l}$.

(b) By making use of the fact that the integral is to be carried out around a closed loop show that an equivalent formula for $\boldsymbol{\tau}$ is

$$\boldsymbol{\tau} = \frac{1}{2} i \left[\oint \mathbf{r} \times d\mathbf{l} \right] \times \mathbf{B}$$

(c) Finally, show that the integral

$$\frac{1}{2} \oint \mathbf{r} \times d\mathbf{l}$$

is the vectorial area of the loop, and thus establish the general validity of (26-24).

***32.** Show that (26-24) is valid for a loop of arbitrary shape by replacing the given current loop by a collection of contiguous, very thin rectangular loops each carrying the same current i.

***33.** Consider a wire of fixed length l

and carrying a current i. This wire can be formed into various loops, such as a square of side $l/4$, N circular loops each of radius $l/2\pi N$, and so on. Show that the maximum torque on any of these in a uniform magnetic induction $\mathbf{B}$ is achieved when the wire forms a circle of radius $l/2\pi$, and calculate the torque in this case.

34. A circular loop of wire of mass m and radius a carries a current i and is free to rotate about a horizontal diameter AC in the presence of a uniform **B**-field directed vertically upward. If it is distributed by a slight amount from its equilibrium position, show that it will oscillate about AC with simple harmonic motion of period

$$P = \left[\frac{2\pi m}{iB}\right]^{1/2}$$

(*Note*: The moment of inertia I of this loop about a diameter is $ma^2/2$.)

35. In an experiment to measure the Hall effect in sodium, suppose that a magnetic induction of strength 0.8 tesla is used and a current of 10 amperes is measured. Assuming that the cross-sectional area of the metallic strip is $2\,\text{cm}^2$ and $E_t = 10^{-5}\,\text{V/m}$, calculate:
   (a) The drift velocity of the electrons.
   (b) The value for the Hall coefficient.
   (c) The number of charge carriers per unit volume and compare this with the number of sodium atoms per unit volume. The atomic mass of sodium is 23 and its density $\cong 10^3\,\text{kg/m}^3$.

*36. Suppose that the pivoted-coil galvanometer in Figure 26-17 has a radial **B**-field of strength 0.3 tesla, and the coil itself has 200 turns and an area of $3.0\,\text{cm}^2$. Calculate the spring constant ($\equiv$ torque/angular displacement), assuming that a current of 1.0 mA produces an angular deflection of 15°.

27 Faraday's law of magnetic induction

What led me . . . to the special theory of relativity was the conviction that the electromotive force acting on a body in motion in a magnetic field was nothing else but an electric field.

ALBERT EINSTEIN (1952)

Some day you will tax it.

MICHAEL FARADAY (On being asked by P. M. Gladstone what was the use of electricity)

27-1 Introduction

If a particle of charge q is at rest relative to an observer, then at a displacement $\mathbf{r}$ from the particle he measures the electric field

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0 r^3}\mathbf{r} \tag{27-1}$$

Consider now this same particle, but this time as seen by an observer relative to whom it travels at the velocity $\mathbf{v}$. In addition to the electric field $\mathbf{E}$ this observer will also detect a $\mathbf{B}$-field given by (see Problem 9, Chapter 25)

$$\mathbf{B} = \frac{\mu_0 q}{4\pi r^3}\mathbf{v} \times \mathbf{r}$$

and which, by use of (27-1), may be cast into the equivalent form

$$\mathbf{B} = \mu_0\epsilon_0\mathbf{v} \times \mathbf{E} \tag{27-2}$$

Bearing in mind the superposition principle and the fact that **v** is the velocity of the particle, we see that associated with an electric field source in motion there is, in general, both a magnetic as well as an electric field.

In 1839 Michael Faraday (1791–1867) first reported on his observations of a related effect, namely, that an *electric field is also, in general, associated with a magnetic source in motion.* That is, even though no electric charge may be present, an electric field comes into existence whenever there is relative motion between a magnetic source and the observer. Moreover, since a **B**-field, which varies in time because the current which produces it does, is physically equivalent to the **B**-field produced by a time-independent magnetic source in motion it follows that an electric field is also associated, in general, with time-dependent magnetic sources. The precise relation between this induced electric field and the properties of the causal, time-varying magnetic source is described quantitatively by a relation known as *Faraday's law.* The purpose of this chapter is to study this important physical law.

27-2 Experimental verification

In this section we describe a number of simple experiments which confirm the fact that an electric field is generally associated with any time-varying **B**-field.

Consider first a loop of wire containing a galvanometer and suppose, as in Figure 27-1a, that the north pole of a permanent magnet approaches the loop. Under this circumstance we find that a current, which is directed as shown, flows in the loop. Similarly, if the north pole recedes from the loop, then as in Figure 27-1b the current flows in the opposite sense. Figures 27-1c and 27-1d show the analogous results obtained by use of the south pole of a

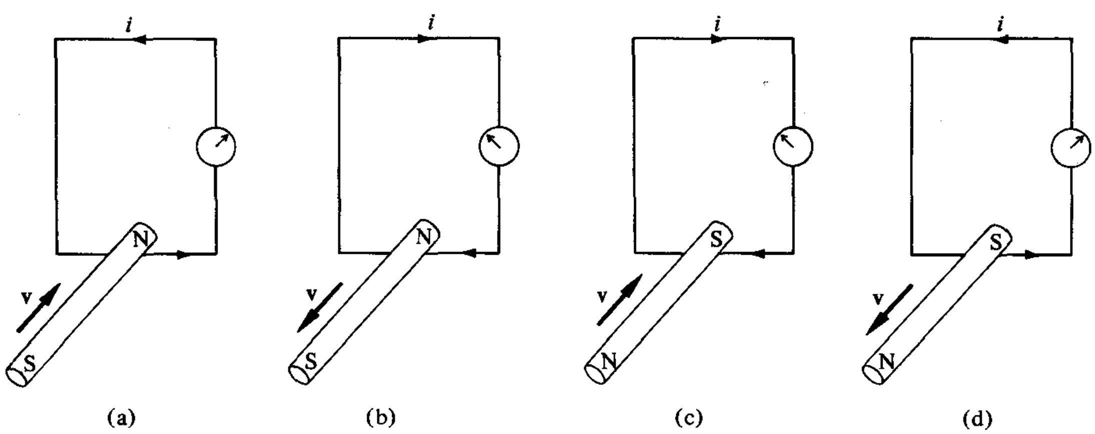


Figure 27-1

permanent magnet. Note that in each case the very existence of current in the loop implies that at least inside the wire there must exist an electric field. Further, experiment shows that if the relative motion between the magnet and the loop ceases, then so does the current. This implies that an electric field is associated only with the *motion* of a magnetic source and not with the source itself.

In connection with experiments such as those in Figure 27-1 there are three features worthy of special emphasis. First, the fact that the magnet is in motion while the loop is stationary in each case is not relevant. Were we to view the situation, say, in Figure 27-1a, from the point of view of an observer moving toward the loop at the velocity **v** (according to whom the magnet is stationary while the loop moves toward the magnet at the velocity −**v**), we would find the current flow in the loop to be unaltered. A second feature is that the use of a permanent magnet is in no way restrictive. If in each of the cases in Figure 27-1 the magnet were replaced by the appropriate end of a coil of wire carrying a current, then precisely the same results would be obtained. In other words, when in motion, a wire carrying a current also produces an electric field. The third feature is that a force must be exerted on the magnets in Figure 27-1 in order to sustain the motion. Reference to the figure shows that the current induced in the loops produces, in each case, a **B**-field, which exerts a force that tends to oppose the motion of the magnet. Thus to sustain the given motion an external agent must exert a force and carry out positive work on the moving magnet.

Having established that an electric field is associated with the motion of magnetic sources, we now describe briefly how to verify the fact that an electric field is also produced by a stationary but *time-dependent* magnetic source. To this end consider, in Figure 27-2, two coplanar circular loops A and B. Suppose that B contains a galvanometer and that a current may be generated in A by closing a switch S. Experiment shows that immediately after S is closed there is a momentary flow of current around B and that this flow ceases shortly afterward. By contrast, the current around loop A becomes steady at the value $\mathscr{E}/R$. Similarly, when the switch S is subsequently opened, so that the current in A drops to zero, there is again a momentary flicker of current around B. This time it flows in the opposite

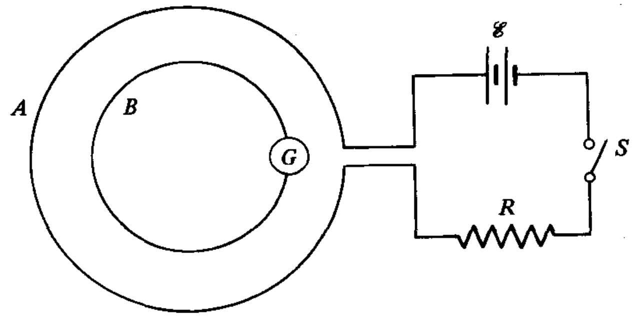


Figure 27-2

direction. It follows that since there is a change of current in A whenever S is opened or closed, an electric field must be associated with a *change* in current. But according to the law of Biot-Savart, a time-varying current will generate a time-varying **B**-field. Hence we conclude that an electric field is associated not only with electric charge and with the motion of magnetic sources, but also with a stationary but time-varying magnetic source.

27-3 The electric field produced by moving magnetic sources

Before proceeding with a study of Faraday's law in its most general form, let us consider first the special case of the electric field associated with moving magnetic sources. It will be assumed throughout this discussion that all velocities are small compared to the speed of light c ($\cong 3 \times 10^8$ m/s). This means that we shall concern ourselves only with those cases in which the relative velocity $\mathbf{v}$ between the observer and the magnetic source is small compared to c. For situations involving speeds $v \sim c$, the theory of relativity is required.

Consider a stationary magnetic source—the north pole of a magnetized rod in Figure 27-3a, for example—and let $\mathbf{B}$ represent the magnetic induction at some fixed point P. Figure 27-3b shows this same system, but this time as seen by an observer with respect to whom the magnetized rod has the velocity $\mathbf{v}_s$ perpendicular to the rod. If this observer carries out a sequence of experiments involving the measurements of forces on charged particles, he would find at point P the same magnetic induction $\mathbf{B}$ as above. In addition, however, he would also find that at point P there exists an electric field $\mathbf{E}$, given by

$$\mathbf{E} = -\mathbf{v}_s \times \mathbf{B} \tag{27-3}$$

Very generally, then: *If the source of a magnetic induction field* $\mathbf{B}$ *is moving at a velocity* $\mathbf{v}_s$ *relative to an observer, then this observer sees the original* **B***-field as well as the electric field* ($-\mathbf{v}_s \times \mathbf{B}$). Note that two observers in relative motion will, in general, observe *different* electric fields.

It is worth emphasizing that this electric field $\mathbf{E}$ in (27-3) as seen by an observer with respect to whom a magnetic source is in motion is not simply a mathematical abstraction; it is a real and measurable electric field. Thus if the observer places a particle of charge q at the point P in Figure 27-3b, he

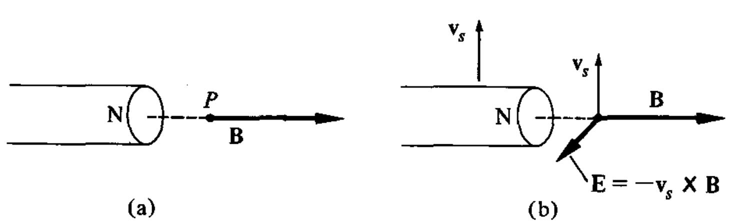


Figure 27-3

measures a force $\mathbf{F} = q\mathbf{E}$ on the particle. This situation is analogous to that in Newtonian mechanics, where in making use of accelerated coordinate systems it is necessary to introduce "fictitious" forces. As we saw there, these forces are very real and directly observable. However, electric forces arise even for magnetic sources in unaccelerated, *uniform* motion. Because of the fact that Newton's laws are invariant under Galilean transformations, no fictitious forces arise for such motions. By contrast, the electric field $\mathbf{E}$ in (27-3) comes into existence with all types of motion of a magnetic source, including uniform motion.

Now even though the validity of (27-3) rests mainly on experimental grounds, the existence of this electric field can also be justified on the basis of logical consistency. To see this, consider in Figure 27-4a a particle of charge q at rest at a point where there is a magnetic induction $\mathbf{B}$ produced, say, by a loop of wire carrying a current i. Since the particle is at rest and $\mathbf{E} = 0$ there is no force on the particle. Hence its acceleration vanishes. Let us now view this same physical situation, but this time as shown in Figure 27-4b as seen by an observer with respect to whom the loop and the particle have an upward velocity $\mathbf{v}_s$. According to this observer, the particle has a velocity $\mathbf{v}_s$ and is in a magnetic induction $\mathbf{B}$. Hence it experiences a magnetic force $\mathbf{F}_m$, given by

$$\mathbf{F}_m = q\mathbf{v}_s \times \mathbf{B}$$

and if there were no other force acting on it there would be a contradiction, since according to Newton's law the particle would accelerate. This seeming paradox is resolved by noting that the particle is also subject to the electric field $\mathbf{E}$ in (27-3). Thus it experiences also an electric force $\mathbf{F}_e$

$$\mathbf{F}_e = q\mathbf{E} = -q\mathbf{v}_s \times \mathbf{B}$$

and it is the sum $(\mathbf{F}_e + \mathbf{F}_m)$ that acts on the particle. On adding these two formulas together, we find that because of (27-3) the total force acting $(\mathbf{F}_m + \mathbf{F}_e)$ is indeed zero and, consistent with our physical expectations, the particle does not accelerate. Hence we see that if a magnetic source is in

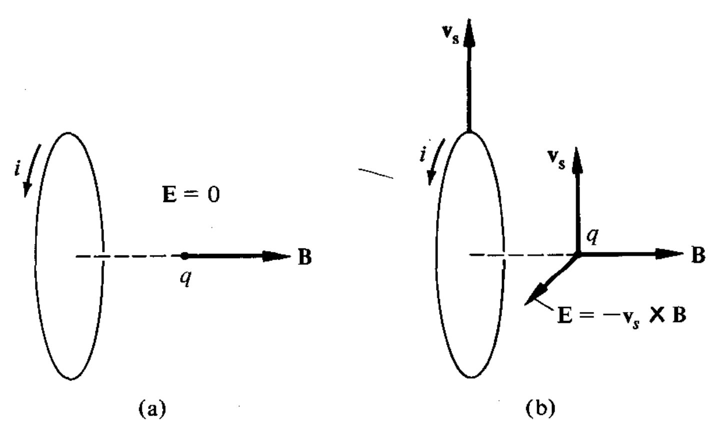


Figure 27-4

motion relative to an observer, an electric field of precisely the form in (27-3) must come into existence in order that inconsistencies such as this do not arise.

Example 27-1 Consider the uniform magnetic induction $\mathbf{B}_0$ of strength 0.02 tesla in a very long ideal solenoid. What is the electric field seen by an observer who travels at a velocity of 10 m/s:

(a) Along the axis of the solenoid?
(b) At right angles to the axis of the solenoid?

Solution

(a) Since the magnetic induction in an ideal solenoid is directed along its axis and since the cross product of two parallel vectors vanishes, it follows from (27-3) that, for this case,

$$\mathbf{E} = -\mathbf{v}_s \times \mathbf{B} = 0$$

Thus, as shown in Figure 27-5a, there are no electric field lines according to this observer.

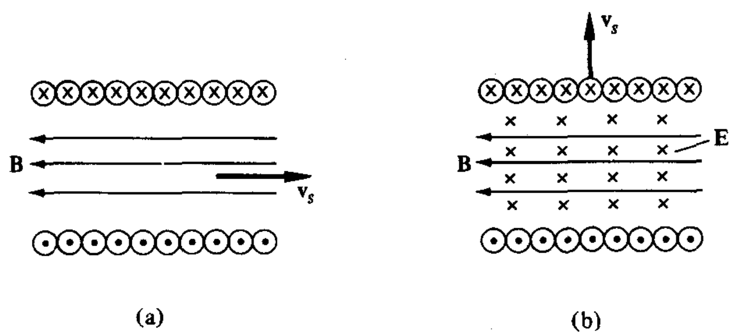


Figure 27-5

(b) Assuming that the observed velocity $\mathbf{v}_s$ of the solenoid is upward (see Figure 27-5b), it follows from (27-3) that the electric field $\mathbf{E}$ is perpendicular to and directed down into the plane of the diagram. The strength of this electric field is

$$\begin{aligned} E = |\mathbf{E}| &= |-\mathbf{v}_s \times \mathbf{B}| = (10\ \text{m/s}) \times (0.02\ \text{T}) \\ &= 0.2\ \text{V/m} \end{aligned}$$

Example 27-2 Consider an infinitely long, straight wire that carries a current i as seen by an observer, with respect to whom the wire travels at a velocity $\mathbf{v}_s$ along the direction of the current. Calculate the electric field according to this observer.

Solution The situation is shown in Figure 27-6. According to (25-6), the strength of the magnetic induction at a distance r from the wire is

$$B = \frac{\mu_0 i}{2\pi r} \tag{27-4}$$

and the $\mathbf{B}$-field lines are circles, concentric with the wire and with a sense given by the right-hand rule. According to the definition of the cross product, the direction of the vector $(-\mathbf{v}_s \times \mathbf{B})$ is everywhere radial. Hence, by use of (27-4) and (27-3), we find

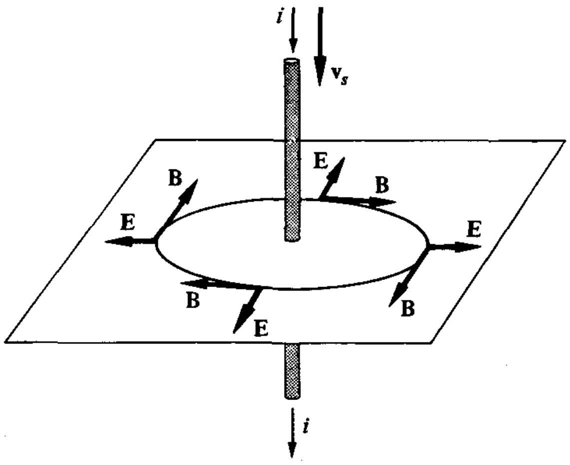


Figure 27-6

that

$$\mathbf{E} = -\mathbf{v}_s \times \mathbf{B} = \mathbf{e}\frac{\mu_0 v_s i}{2\pi r} \tag{27-5}$$

with $\mathbf{e}$ a unit vector in the radial direction.

According to this observer, in addition to the **B**-field there is also the radial electric field given in (27-5). Since the electric field associated with an infinite line charge, of charge per unit length λ, is $\mathbf{e}\lambda/2\pi\epsilon_0 r$ (see (20-6)), it follows on comparison with (27-5) that, in addition to the current i, this observer also sees on the wire a charge per unit length

$$\lambda = \mu_0 \epsilon_0 v_s i \tag{27-6}$$

Note that this charge density λ and its associated electric field are very real and directly observable. Assuming the values $v_s = 10^3$ m/s and $i = 10$ amperes, we find by use of (27-6) that

$$\begin{aligned}\lambda &= \mu_0 \epsilon_0 v_s i \\ &= (4\pi \times 10^{-7}\ \text{T-m/A}) \times (8.85 \times 10^{-12}\ \text{F/m}) \times (10^3\ \text{m/s}) \times (10\ \text{A}) \\ &= 1.1 \times 10^{-13}\ \text{C/m}\end{aligned}$$

27-4 The motion of a conductor in a magnetic field

One of the very important applications of the electric field formula in (27-3) is to the problem of a conductor in motion relative to a magnetic source. The purpose of this section is to describe certain physical effects associated with this motion.

First, however, let us consider the case of a thin, conducting rod of length l traveling in a uniform **B**-field at a velocity $\mathbf{v}$ perpendicular to its axis. See Figure 27-7a. From the viewpoint of an observer at rest relative to the rod, the magnetic sources are traveling at the velocity $\mathbf{v}_s = -\mathbf{v}$. Hence, inside the

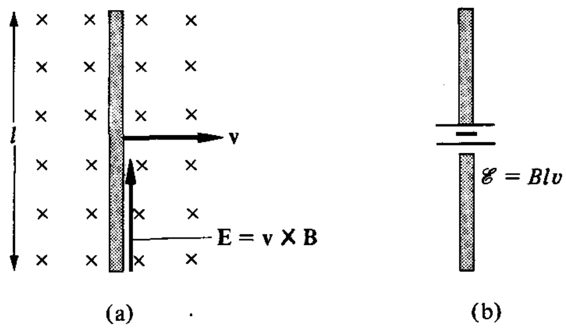


Figure 27-7

conductor there is an electric field, given in accordance with (27-3) by

$$\mathbf{E} = \mathbf{v} \times \mathbf{B} \tag{27-7}$$

(Note the sign!) As shown in Figure 27-7a, for the given directions of **B** and **v** this electric field is directed vertically upward along the rod and has a strength vB. In effect, then, there exists a uniform electric field of strength Bv inside the wire. But this electric field cannot be sustained! For the electrons inside the rod will respond to the field and rearrange themselves in a way so as to nullify its effects. Thus positive and negative charge will accumulate at the upper and lower end of the rod, respectively, until eventually the total electric field inside the rod vanishes. In effect then, as shown in Figure 27-7b, the ends of the rod become the two terminals of a seat of electromotive force with emf $\mathscr{E}$ given by

$$\mathscr{E} = Blv \tag{27-8}$$

and with the upper end of the rod the positive terminal. Note that since the rod is not connected to a closed conducting path, no current flows.

Alternatively, we may think of this redistribution of charge inside the rod in the following way. The free electrons in its interior share in the rod's motion and thus travel at the velocity **v** directed to the right in Figure 27-7a. Because of their negative charge $(-q)$, the force on these electrons $(-q\mathbf{v} \times \mathbf{B})$ is directed downward along the rod. Hence, just as above, negative charge accumulates at the lower end of the rod and a compensating positive charge appears at the other end.

Let us now reexamine this situation, but this time suppose as shown in Figure 27-8 that the moving conductor slides on two metallic rails and is thereby part of a closed conducting path. For simplicity assume that the moving conductor has negligible resistance while the rest of the current path has a constant resistance R. Since only the rod is in motion relative to the magnetic sources, it alone experiences the electric field in (27-7) and the emf in (27-8) is developed only across it. In effect, then, we have the situation shown in Figure 27-8b, of a battery of emf $\mathscr{E} = Blv$ connected across a resistor R. According to Ohm's law, the current i is

$$i = \frac{\mathscr{E}}{R} = \frac{Blv}{R} \tag{27-9}$$

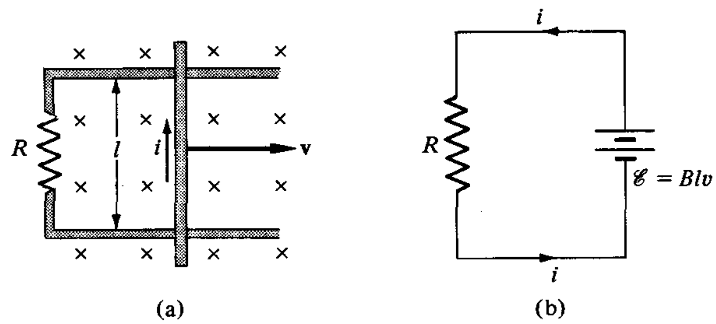


Figure 27-8

and the direction of this current is upward along the moving conductor in Figure 27-8a. According to (26-15), the external **B**-field exerts on the current in the moving rod a force **F**, which is directed to the left and has the magnitude

$$|\mathbf{F}| = ilB = \frac{l^2B^2v}{R} \tag{27-10}$$

where the second equality follows by use of (27-9). To keep the rod moving to the right at the given velocity **v** it is necessary to exert an external force of strength given in (27-10) but directed to the right.

Figure 27-9 illustrates the fact that current need not flow in a conductor moving through a **B**-field under all circumstances. If a rectangular conducting loop travels at some velocity **v** through a uniform **B**-field, directed as shown, then no current flows. This time an emf of strength Blv directed upward is developed in each vertical segment. Hence this situation is physically the same as that shown in Figure 27-9b, in which two identical batteries are connected in a way so that no current flows. By contrast, if the loop were to travel through an inhomogeneous **B**-field, say that in Figure 27-10a, with only one of the vertical wire segments in the **B**-field, then current would again flow. Figure 27-10b shows the equivalent circuit diagram for this case.

Figure 27-11 shows the general case of a thin loop of wire traveling at a velocity **v** through a **B**-field, with **v** and **B** at an arbitrary angle. If **B** is the magnetic induction at the position of an element $d\mathbf{l}$ of the loop, then according to (27-7) the electric field **E** in this element is $\mathbf{v} \times \mathbf{B}$. Hence the emf $d\mathscr{E}$

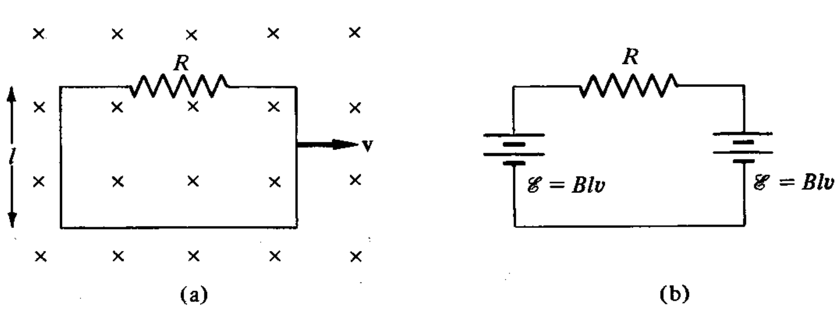


Figure 27-9

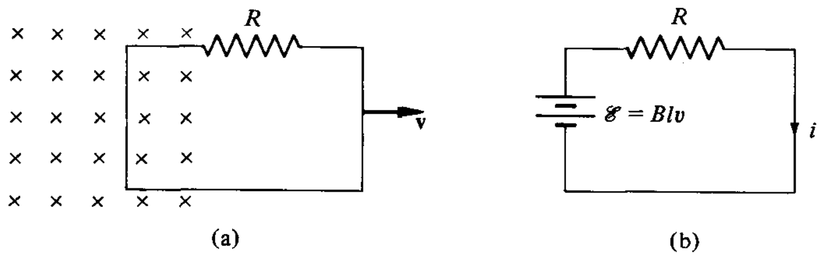


Figure 27-10

developed between its endpoints is (Section 23-2)

$$d\mathscr{E} = \mathbf{E} \cdot d\mathbf{l} = (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} \tag{27-11}$$

and the emf $\mathscr{E}$ developed around the entire loop is obtained by adding together the contributions from all elements of the loop. The result is

$$\mathscr{E} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} \tag{27-12}$$

where the line integral is to be carried out over the entire loop. In general, different parts of the loop may travel at different speeds, and the magnetic induction may not be uniform around the loop. All such variations must be allowed for in evaluating the line integral in (27-12).

Example 27-3 If, in Figure 27-8, $B = 0.1$ tesla, $l = 10$ cm, $v = 15$ m/s, and $R = 10\ \Omega$, calculate:

(a) The emf developed between the ends of the rod.
(b) The current.
(c) The force required to keep the rod moving.

Solution

(a) Substitution into (27-8) yields

$$\mathscr{E} = Blv = (0.1\ \mathrm{T}) \times (0.1\ \mathrm{m}) \times (15\ \mathrm{m/s}) = 0.15\ \mathrm{V}$$

(b) The current is given by Ohm's law:

$$i = \frac{\mathscr{E}}{R} = \frac{0.15\ \mathrm{V}}{10\ \Omega} = 1.5 \times 10^{-2}\ \mathrm{A}$$

(c) The force required to keep the rod moving is, according to (27-10),

$$F = \frac{l^2B^2v}{R} = \frac{(0.1\ \mathrm{m})^2 \times (0.1\ \mathrm{T})^2 \times (15\ \mathrm{m/s})}{10\ \Omega}$$
$$= 1.5 \times 10^{-4}\ \mathrm{N}$$

The direction of this force is parallel to **v**.

Example 27-4 Consider again the situation in Figure 27-8. Show that the rate at which the external agent must expend energy to keep the conductor traveling at the velocity **v** is the same as the power dissipated in the resistor.

Solution The rate dW/dt at which the external agent carries out work is

$$\frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v}$$

where $\mathbf{F}$ is the force exerted. Substituting the value for $|\mathbf{F}|$ in (27-10) and noting that $\mathbf{F}$ and $\mathbf{v}$ are parallel, we find that

$$\frac{dW}{dt} = \frac{l^2 B^2 v^2}{R} \tag{27-13}$$

On the other hand, according to (27-9), the current i in the resistor is

$$i = \frac{Blv}{R}$$

and substitution into (27-13) yields

$$\frac{dW}{dt} = Ri^2$$

Finally, since Ri^2 is the power dissipated in the resistor, the desired result follows.

Example 27-5 A wire of length l is forced to travel at a velocity $\mathbf{v}$ along the direction of the current in a long, straight wire (see Figure 27-12). Calculate the emf developed between the ends of the moving wire.

Solution Let $d\mathscr{E}$ represent the emf developed across an element of length dx of the moving wire at a distance x from the current. Making use of (27-11) and the fact that the magnetic induction $\mathbf{B}$ at this element is directed vertically down into the plane of the figure and has the strength $B = \mu_0 i/2\pi x$, we find that

$$d\mathscr{E} = -v\frac{\mu_0 i}{2\pi x}\, dx \tag{27-14}$$

The minus sign reflects the fact that the electric field in the moving conductor is directed toward the current. Integrating over all values from $x = a$ to $x = (a + l)$ we

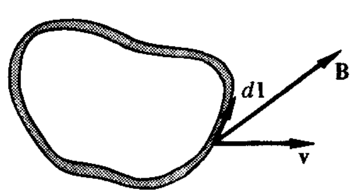


Figure 27-11

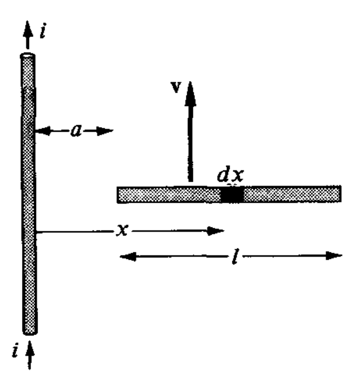


Figure 27-12

find for the emf across the wire

$$\mathscr{E} = \int d\mathscr{E} = \int_a^{a+l} \left(-\frac{v\mu_0 i\, dx}{2\pi x}\right)$$
$$= -\frac{v\mu_0 i}{2\pi}\int_a^{a+l}\frac{dx}{x} = -\frac{v\mu_0 i}{2\pi}\ln x\Big|_a^{a+l}$$
$$= -\frac{v\mu_0 i}{2\pi}\ln\left(1+\frac{l}{a}\right)$$

Again, the minus sign reflects the fact that the end of the moving rod closest to the current is at the higher potential.

Example 27-6 A rectangular loop of wire of sides a and l is situated near a long, straight wire in which there flows a current i. Suppose, as shown in Figure 27-13, that the loop travels to the right with a velocity $\mathbf{v}$. Calculate the emf developed around the loop.

Solution Since the magnetic induction $\mathbf{B}$ on the left segment of the loop has the magnitude $\mu_0 i/2\pi x$ and is directed down into the plane of the figure, it follows (see (27-8)) that the emf $\mathscr{E}_L$ on this segment is directed upward and has the magnitude

$$\mathscr{E}_L = \frac{vl\mu_0 i}{2\pi x}$$

Similarly, the emf $\mathscr{E}_R$ on the right-hand segment is also directed upward and has the value

$$\mathscr{E}_R = \frac{vl\mu_0 i}{2\pi(x+a)}$$

Since the induced electric field $\mathbf{v}\times\mathbf{B}$ has no components along the upper and the lower segments of the loop, no emf is developed along these. Hence the total emf $\mathscr{E}$ around the loop is clockwise and has the value

$$\mathscr{E} = \mathscr{E}_L - \mathscr{E}_R = \frac{\mu_0 ilv}{2\pi}\left(\frac{1}{x} - \frac{1}{x+a}\right)$$

Thus we see that if a closed loop travels through an *inhomogeneous* **B**-field an emf is developed around it. This is to be contrasted with the fact, established in the problems and exemplified in Figure 27-9, that no emf is developed around a rigid loop moving through a *uniform* **B**-field.

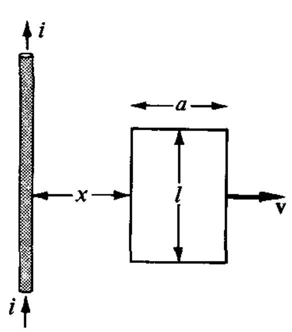


Figure 27-13

27-5 Lenz's law for moving conductors

In Example 27-4 it was established that the current through the moving conductor in Figure 27-8 is directed so that the external force that causes the conductor to move carries out positive work. In other words, the current is directed so that it tends to oppose the relative motion of the conductor and the magnetic sources. In 1834, Heinrich Lenz (1804–1865) showed that this property is very generally true. In his own words:

> If a constant current flows in the primary circuit A and if by the motion of A, or of the secondary circuit B, a current is induced in B, the direction of this induced current will be such that by its electromagnetic action on A it tends to oppose the relative motion of the two circuits.

The purpose of this section is to derive this law.

Suppose that the rigid loop of wire in Figure 27-11 has a resistance R and is forced to travel at a velocity $\mathbf{v}$ through a magnetic induction $\mathbf{B}$ under the action of an external force $\mathbf{F}$. In order to establish the validity of Lenz's law, it is necessary to prove that this force carries out positive work in this process. We shall now show that the rate $\mathbf{F} \cdot \mathbf{v}$ at which the external force carries out work is invariably positive. Lenz's law then follows as a direct consequence.

The emf $\mathscr{E}$ developed around the moving loop in Figure 27-11 is given by (27-12). Making use of the vector identity

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$$

which is valid for any three vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, this formula for $\mathscr{E}$ may be expressed equivalently as

$$\mathscr{E} = \oint \mathbf{v} \cdot (\mathbf{B} \times d\mathbf{l}) = \mathbf{v} \cdot \oint \mathbf{B} \times d\mathbf{l} \qquad \text{(27-15)}$$

where the second equality follows since the loop is presumed to be rigid and hence each of its elements $d\mathbf{l}$ has the same velocity.

Now the force $\mathbf{F}_B$ *on* the loop due to the external $\mathbf{B}$-field is

$$\mathbf{F}_B = \oint i\, d\mathbf{l} \times \mathbf{B} = -i \oint \mathbf{B} \times d\mathbf{l} \qquad \text{(27-16)}$$

where the second equality follows from the vector identity in (10-3). On the other hand, the force $\mathbf{F}$ that the external agent must exert to keep the loop moving at the velocity $\mathbf{v}$ is equal and opposite to $\mathbf{F}_B$. Hence the rate $\mathbf{v} \cdot \mathbf{F}$ at which the external agent carries out work is

$$\begin{aligned} \mathbf{v} \cdot \mathbf{F} &= -\mathbf{v} \cdot \mathbf{F}_B = -\mathbf{v} \cdot \left[-i \oint \mathbf{B} \times d\mathbf{l} \right] \\ &= i\mathbf{v} \cdot \oint \mathbf{B} \times d\mathbf{l} = i\mathscr{E} \qquad \text{(27-17)} \\ &= Ri^2 > 0 \end{aligned}$$

where the second equality follows from (27-16), the fourth from (27-15), and the fifth by use of Ohm's law. Since the quantity Ri^2 is inherently positive, it follows that so is the rate $\mathbf{v} \cdot \mathbf{F}$ at which work is carried out by the external agent. The validity of Lenz's law is thereby established.

It is interesting to note that, according to (27-17), the rate at which work is carried out by the external agent is the same as that at which heat is dissipated in the resistor. Thus, consistent with the ideas of energy conservation, the work carried out by the external agent is converted entirely to another form of energy.

Example 27-7 A small loop of wire has a resistance of 10 Ω and a force $\mathbf{F}$ of 0.1 newton is required to pull it at a velocity of 2.0 m/s through an inhomogeneous magnetic induction. Assuming that the direction of $\mathbf{F}$ is parallel to that of $\mathbf{v}$:

(a) What is the rate at which the external agent carries out work?
(b) What current flows around the loop?
(c) What emf is developed around the loop?

Solution

(a) Using the given values for $\mathbf{F}$ and $\mathbf{v}$ we find that

$$\mathbf{v} \cdot \mathbf{F} = (2.0\ \text{m/s}) \times (0.1\ \text{N}) = 0.2\ \text{W}$$

(b) On equating this power to the ohmic heating losses in accordance with (27-17) we obtain

$$0.2\ \text{W} = Ri^2 = (10\ \Omega)i^2$$

and this yields

$$i = 0.14\ \text{A}$$

(c) Making use of these values for R and i, we find by use of Ohm's law that

$$\mathscr{E} = Ri = (10\ \Omega) \times (0.14\ \text{A}) = 1.4\ \text{V}$$

27-6 Faraday's law

So far in this chapter we have considered only the electric field—and the associated emf in a conductor—due to the relative motion between an observer and magnetic sources. As noted previously, a $\mathbf{B}$-field that varies in time due to such a motion is physically indistinguishable from one that varies in time for other reasons, for example, because the currents in the sources themselves vary in time. We might expect, therefore, that an electric field is very generally associated with *all* time-varying magnetic fields. This is indeed the case. The precise form of this relationship between a $\mathbf{B}$-field that varies in time for any reason and its associated electric field is given by *Faraday's law.* The purpose of this section is to describe this all-important law.

Consider, in Figure 27-14, a region of space in which there is a time-varying $\mathbf{B}$-field and let S be an arbitrary open surface that bounds a given closed curve l. As in the discussion of Ampère's law (Chapter 25), let us

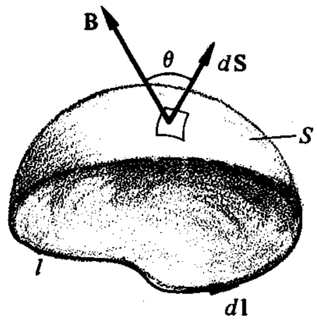


Figure 27-14

assign a sense of direction along the curve and determine a positive sense to the normal to S in accordance with the right-hand rule. (If the curve is grasped in the right hand with the fingers pointing along the sense of l, then the outstretched thumb will point along the positive sense of the normal to S.) The magnetic flux Φ_m through the surface S is defined by

$$\Phi_m = \int_S B \cos\theta \, dS = \int_S \mathbf{B} \cdot d\mathbf{S} \tag{27-18}$$

with $d\mathbf{S}$ a vectorial area element along the normal to S and with θ the angle between $\mathbf{B}$ and $d\mathbf{S}$. Now according to Gauss' law for magnetism, (25-11), the total magnetic flux out of every closed surface vanishes. Hence the magnetic flux Φ_m in (27-18) is the same for *all* open surfaces S bounded by the given curve l. In other words, the magnetic flux Φ_m is determined exclusively by the bounding curve itself and not by the particular surface S chosen.

Consider now a region of space in which there exist an electric field $\mathbf{E}$ and a magnetic induction field $\mathbf{B}$. Faraday's law states that for *all* closed curves l, $\mathbf{E}$ and $\mathbf{B}$ are related by

$$\oint_l \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S} \tag{27-19}$$

where the integral on the left is to be carried out around the closed curve l, and on the right the surface integral is over any open surface bounded by l. It is assumed that the positive sense of l and S are related as in Figure 27-14 by the right-hand rule. Implicit in (27-19) is the fact that this relation is valid for *all* closed curves l.

An alternate way of expressing Faraday's law is in terms of magnetic flux. Substituting (27-18) on the right-hand side of (27-19) we obtain

$$\oint_l \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \Phi_m \tag{27-20}$$

which states in words:

The line integral of the electric field about any closed path l is equal to the negative of the time rate of change of magnetic flux through any surface bounded by l.

For the special, but frequently occurring, case where the bounding curve l is selected to coincide with a loop of a conducting wire, the line integral in (27-20) can be identified with the emf $\mathscr{E}$ developed around it. Faraday's law in this case assumes the form

$$\mathscr{E} = -\frac{d}{dt}\Phi_m \tag{27-21}$$

or, in words,

The emf developed around a closed conducting path is equal to the negative of the rate of change of magnetic flux through any surface bounded by that path.

In particular, the current around a closed loop in a time-varying **B**-field is the same as if a battery of emf $(-d\Phi_m/dt)$ were inserted in the loop.

It should be noted that just as Ampère's law in (25-12) does not by itself determine a unique **B**-field for a given current distribution, by the same token Faraday's law in (27-19) does not associate a unique electric field with a given time-dependent **B**-field. Nevertheless, Faraday's law is very useful in a practical sense and is one of the very important characterizations of the electric field. In particular, as noted above, if the closed loop l in (27-19) is selected to coincide with a closed conducting path, then the line integral in that relation represents the emf developed around this conductor. In many cases of practical importance, only this emf is of direct physical interest. And for these cases, according to (27-21), the significant quantity is the rate of change of magnetic flux through the conducting loop; the details of the electric field itself are at best only of peripheral interest.

As an illustration, consider again the loop of wire traveling through a uniform **B**-field in Figure 27-9. Because **B** is uniform, the magnetic flux through the loop does not vary in time. Hence, consistent with our previous analysis the emf around the loop vanishes. By the same token, in Figure 27-10, the loop travels through an inhomogeneous field and thus the magnetic flux through it varies in time. Hence current flows around the loop in this case.

Example 27-8 A circular conducting loop of radius a lies in a plane at right angles to a time-dependent but uniform magnetic induction, whose strength is

$$B(t) = \alpha + \beta t$$

where α and β are positive constants (see Figure 27-15). Calculate the emf developed around the loop.

Solution Assuming the sense for the bounding curve shown in the figure, the flux Φ_m through the planar surface bounded by the loop is

$$\Phi_m = \int_S \mathbf{B} \cdot d\mathbf{S} = B\pi a^2 = (\alpha + \beta t)\pi a^2$$

Substituting this result into (27-21) we find that

$$\mathscr{E} = -\frac{d\Phi_m}{dt} = -\frac{d}{dt}[(\alpha + \beta t)\pi a^2] = -\beta\pi a^2$$

The fact that this is negative (since $\beta > 0$) means that the emf developed in the loop is opposite to the sense assumed in Figure 27-15. This implies, in turn, that the resultant current in the loop is directed counterclockwise. If the flux had been decreasing, that is, if $\beta < 0$, then the current would have been directed clockwise.

Example 27-9 Consider, in Figure 27-16, the uniform but time-dependent magnetic induction $\mathbf{B}(t)$ between the pole faces of an electromagnet. Calculate the strength of the electric field $\mathbf{E}(t)$ in this region. Assume that the electric field lines are circles concentric with the axis of the electromagnet and that the magnetic induction is uniform.

Solution Let us apply Faraday's law to a circular loop of radius r with its center on the axis of symmetry of the electromagnet. Assuming that the sense of the bounding curve is as shown in the figure, we find

$$\oint \mathbf{E} \cdot d\mathbf{l} = 2\pi r E \tag{27-22}$$

with E the strength of the electric field. Since the sense of the bounding curve is such that the normal to the surface bounded by it is downward, the magnetic flux through it is positive and given by

$$\Phi_m = \int \mathbf{B} \cdot d\mathbf{S} = B(t) \int dS = \pi r^2 B(t) \tag{27-23}$$

where the second equality follows since the magnetic induction is uniform.

Substituting (27-22) and (27-23) into Faraday's law, we find that

$$2\pi r E(t) = -\frac{d}{dt}[\pi r^2 B(t)] = -\pi r^2 \frac{dB}{dt}$$

or, in other words,

$$E = -\frac{r}{2}\frac{dB}{dt} \tag{27-24}$$

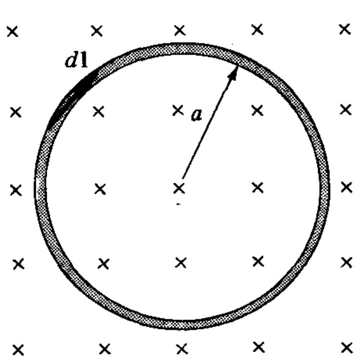


Figure 27-15

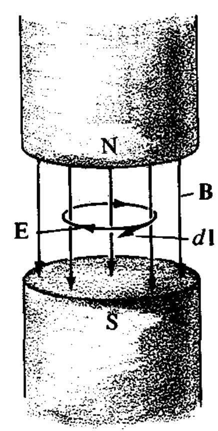


Figure 27-16

Thus, if the field decreases, so that $dB/dt < 0$, then the sense of the electric field lines is as shown in Figure 27-16. However, if the field increases, so that $dB/dt > 0$, then the electric field lines are oriented in the opposite sense.

The physical principle illustrated here is used in a practical way to produce very energetic electrons in a device called a *betatron.* The betatron was invented by Donald W. Kerst at the University of Illinois in 1941 and has been used to produce electron beams with energies as high as 300 MeV. If an electron is introduced into the region between the pole pieces in Figure 27-16, then because of the **B**-field it will orbit (approximately) in a circle, and because of the associated electric field in (27-24) it will also undergo a tangential acceleration. In this way its energy steadily rises to the desired value.

27-7 Applications of Faraday's law

In this section the general utility of Faraday's law in the form in (27-21) will be illustrated by applying it to a number of idealized physical situations.

Consider, in Figure 27-17a, a loop of wire in the presence of a time-dependent external magnetic induction $\mathbf{B}_0(t)$. If the vector $d\mathbf{l}$ represents the sense of the curve, it follows that the magnetic flux through the loop and associated with $\mathbf{B}_0$ is positive. Hence if $\mathbf{B}_0(t)$ is increasing, then $d\Phi_m/dt$ will be positive. The emf in the loop will then be negative, according to (27-21), and the current thus flows in a direction opposite to the sense of $d\mathbf{l}$. By contrast, if as shown in Figure 27-17b the external field $\mathbf{B}_0(t)$ is decreasing, then $d\Phi_m/dt$ will be negative. This time, then, the induced emf is positive, and the current will flow along $d\mathbf{l}$. Note that if $\mathbf{B}_i$ is the **B**-field due to the induced current i in the loop, then for $\mathbf{B}_0$ increasing, $\mathbf{B}_i$ is antiparallel to $\mathbf{B}_0$, and for $\mathbf{B}_0$ decreasing it is parallel to $\mathbf{B}_0$. In either case, the direction of the induced current is such as to produce a **B**-field that tends to *oppose* the *change* in magnetic flux through the loop. This leads to the alternate form of Lenz's law:

> *The direction of the induced current in a loop of wire in the presence of a time-dependent* **B**-*field is such as to oppose changes in magnetic flux through it.*

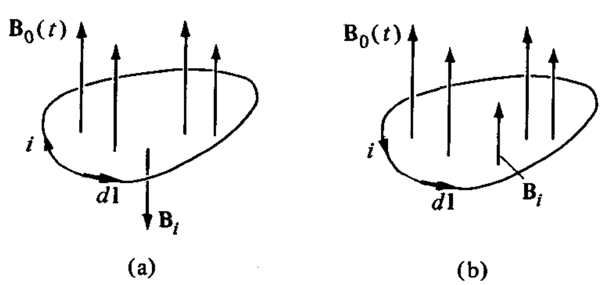


Figure 27-17

Brief reflection shows that this form for Lenz's law is more general than that for moving conductors as derived in Section 27-5.

In working problems involving the application of (27-21) to particular physical situations it is often simplest to disregard the sign in that relation and to use it only to obtain the magnitude of the induced emf. The direction of the associated current flow can then be obtained by use of the above form of Lenz's law.

Example 27-10 Suppose the external field in Figure 27-15 varies in time as

$$B_0 = \alpha + \beta t$$

Determine the direction of the induced current in terms of the parameter β, assuming that $\alpha > 0$.

Solution For the assumed direction of $d\mathbf{l}$ in the figure, the flux through the loop will be positive (at $t = 0$) and an increasing function of time for $\beta > 0$. For this case, then, the induced field must be (perpendicular to and) out of the plane of the figure in accordance with Lenz's law. By the right-hand rule, then the current will be flowing counterclockwise, that is, in a direction opposite to $d\mathbf{l}$.

For the case $\beta < 0$, the flux will be decreasing. Here the induced current must flow clockwise, so the induced field will be parallel to the external field.

Example 27-11 Consider again a conducting rod sliding across two rails in a time-independent and uniform magnetic induction. See Figure 27-18. This time calculate the induced emf by use of (27-21).

Solution First, with regard to the direction of the induced emf and thus that of the current, we may reason as follows. Assuming a clockwise sense for the conducting path, we find that as the conducting rod moves to the right, the magnetic flux through the loop increases, since the area of the loop does. According to Lenz's law, then, the current in the loop will flow counterclockwise so that its associated magnetic induction will be directed opposite to that of the original magnetic induction $\mathbf{B}_0$.

With regard to the magnitude of the induced emf, we proceed as follows. The magnetic flux through the planar area bounded by the loop is $B_0 lx$, since the length and width of the loop are l and x, respectively. Substitution into (27-21) leads to

$$\mathscr{E} = -\frac{d\Phi_m}{dt} = -\frac{d}{dt}(B_0 lx)$$

$$= -B_0 l \frac{dx}{dt} = -B_0 lv$$

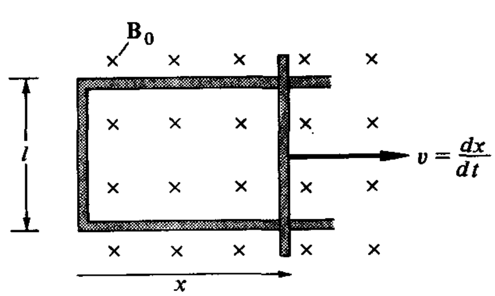


Figure 27-18

where the third equality follows since B_0 and l are constants, and the last since $v = dx/dt$. Note that except for the sign, which is due to our choice for the sense of the bounding curve, this result is identical to that in (27-8).

As this example illustrates, Faraday's law in the form in (27-21) can be utilized not only for the solution of problems involving explicitly time-dependent magnetic sources, but also for those involving magnetic sources in motion. On the other hand, the electric field formula in (27-3) is *not* generally applicable for stationary, but time-dependent, sources. It can be used only for the cases involving *static* magnetic sources in motion.

Example 27-12 A time-dependent current $i = i_0 \sin \omega t$ (i_0 and ω are constants) flows in a long, straight wire which lies in the plane of a nearby rectangular loop (see Figure 27-19). Calculate the emf induced in the loop.

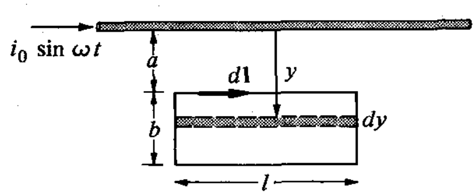


Figure 27-19

Solution As shown, let us assign a clockwise sense to the loop. Then at an instant when the current flows to the right, the magnetic flux through the loop will be positive. It follows from Lenz's law that when the current in the long wire is increasing, the induced current will flow counterclockwise and conversely.

To calculate the emf, let $d\Phi_m$ be the flux through the infinitesimal area element of length l and width dy, and at a distance y below the long wire. Making use of (25-6), we find that

$$\begin{aligned} d\Phi_m &= \mathbf{B} \cdot d\mathbf{S} = Bl\,dy \\ &= \frac{\mu_0 l}{2\pi y}(i_0 \sin \omega t)\,dy \end{aligned}$$

which, when integrated over all values of y from a to $(a+b)$, becomes

$$\begin{aligned} \Phi_m = \int d\Phi_m &= \frac{\mu_0 l i_0 \sin \omega t}{2\pi} \int_a^{a+b} \frac{dy}{y} \\ &= \frac{\mu_0 l i_0 \sin \omega t}{2\pi} \ln y \Big|_a^{a+b} \\ &= \frac{\mu_0 l i_0 \sin \omega t}{2\pi} \ln\left(1 + \frac{b}{a}\right) \end{aligned}$$

Finally, substitution into (27-21) yields the emf

$$\begin{aligned} \mathscr{E} = -\frac{d\Phi_m}{dt} &= -\frac{\mu_0 l i_0}{2\pi} \ln\left(1 + \frac{b}{a}\right) \frac{d}{dt} \sin \omega t \\ &= -\frac{\mu_0 l i_0 \omega}{2\pi} \ln\left(1 + \frac{b}{a}\right) \cos \omega t \end{aligned}$$

The minus sign tells us that for times t for which $\cos \omega t > 0$ the direction of the current in the loop is counterclockwise; that is, opposite to the original positive sense assigned to the loop. For times t for which $\cos \omega t < 0$ the current flows in the opposite direction.

27-8 Alternating current generators

An important problem of considerable practical interest deals with the direct conversion of mechanical to electrical energy. The purpose of this section is to discuss the principles underlying the operation of a device known as an *alternating current generator* which can be used for this purpose.

Consider, in Figure 27-20, a planar coil of wire of area A, which is being rotated about an axis in its plane while in a uniform magnetic induction $\mathbf{B}$. Let us set up a stationary coordinate system with the axis of rotation along the z-axis and with $\mathbf{B}$ directed along the y-axis. If θ is the angle between the normal $\mathbf{n}$ to the plane of the loop and the direction of $\mathbf{B}$, then as the loop rotates at some angular velocity ω, the angle θ increases linearly in time as $\theta = \omega t$. As a result of the rotation, the magnetic flux through the loop varies in time, and therefore, according to Faraday's law, an emf is developed around it. In terms of Figure 27-20, this means that a potential difference will be developed between the two terminals a and b and thus if an electric network is connected across these terminals a current will flow.

To calculate the magnitude of the emf, we proceed as follows. Since $\mathbf{B}$ is constant and spatially uniform, the magnetic flux Φ_m through the loop at the instant shown in the figure is

$$\begin{aligned}\Phi_m &= \int \mathbf{B} \cdot d\mathbf{S} = \int B \cos \theta \, dS \\ &= BA \cos \omega t\end{aligned}$$

where the second equality follows by use of the definition of the dot product,

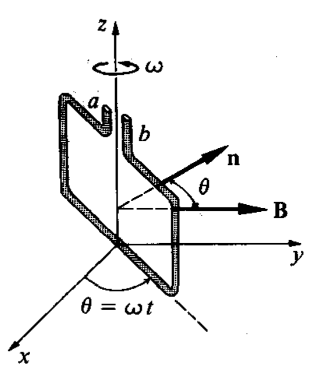


Figure 27-20

and the third since $\theta = \omega t$. Substitution into (27-21) leads to

$$\mathscr{E} = -\frac{d\Phi_m}{dt} = -\frac{d}{dt}(BA\cos\omega t) = -BA\frac{d}{dt}\cos\omega t = \omega BA\sin\omega t \tag{27-25}$$

As long as the loop continues to rotate, this emf will appear across the terminals a and b.

If there are N turns of wire in the loop, then, following the same steps as above, we find the more general formula

$$\mathscr{E} = N\omega BA\sin\omega t \tag{27-26}$$

As a practical matter, the emf developed in a rotating loop can be utilized in an external circuit by use of two *slip rings*, which rotate with the loop (see Figure 27-21). As the slip rings rotate, electrical contact is made with the external circuit by means of two stationary brushes that are permanently in contact with the slip rings.

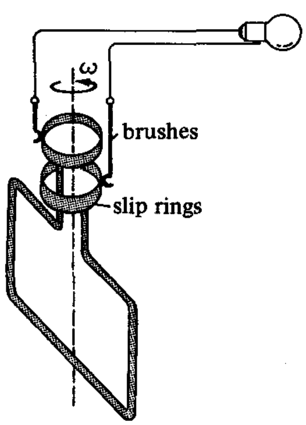


Figure 27-21

When connected to a circuit so that a current flows, the external **B**-field exerts a torque on the rotating loop. According to Lenz's law the direction of this torque is such as to oppose the rotation. Hence an external agent must carry out work to keep the loop rotating.

Example 27-13 Suppose that a resistor of resistance R is connected across the terminals in Figure 27-20. If the coil has N turns each of area A, the magnetic induction is B, and ω is the angular velocity calculate the torque required to keep the loop rotating.

Solution Since the emf $\mathscr{E}$ across the resistor is given according to (27-26) by

$$\mathscr{E} = \omega NBA\sin\omega t$$

it follows from Ohm's law that the current i through the resistor is

$$i = \frac{\mathscr{E}}{R} = \frac{\omega NBA}{R}\sin\omega t$$

Making use of (26-22) which gives the torque on a loop carrying a current i, we find that

$$\tau = iNAB \sin\theta$$
$$= \frac{\omega(NAB)^2}{R} \sin^2 \omega t$$

where the first equality is obtained by making the substitution $\alpha \to \theta$. Note that, in accordance with Lenz's law, this torque is always nonnegative, thus reflecting the fact that the external agent who causes the loop to rotate must carry out positive work during the entire cycle. Note also that the rate at which the external agent carries out work, $\tau\omega$, is precisely the same as the rate Ri^2 at which heat is dissipated in the resistor. Hence all of the mechanical work is converted directly to heat in this case.

27-9 Summary of important formulas

Suppose that **B** is the magnetic induction associated with certain magnetic sources. Then from the viewpoint of an observer with respect to whom these magnetic sources have the velocity $\mathbf{v}_s$ there is, in addition to **B**, an electric field **E**, which at each point in space is

$$\mathbf{E} = -\mathbf{v}_s \times \mathbf{B} \tag{27-3}$$

The emf $d\mathscr{E}$ developed between the ends of a thin, conducting element $d\mathbf{l}$ traveling at a velocity **v** is

$$d\mathscr{E} = \mathbf{E} \cdot d\mathbf{l} = (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{l} \tag{27-11}$$

where **B** is the external (time-independent) **B**-field.

Faraday's law states that for *all* closed curves l and regardless of what electric or magnetic sources may be present, **E** and **B** are related by

$$\oint_l \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt}\int_S \mathbf{B} \cdot d\mathbf{S} \tag{27-19}$$

where S is any open surface bounded by l. (See Figure 27-14 for the relation between the normal to S and sense of the curve l.) For the special case that l is selected to coincide with a loop of wire, (27-19) assumes the form

$$\mathscr{E} = -\frac{d\Phi_m}{dt} \tag{27-21}$$

where $\mathscr{E}$ is the emf developed around the loop and Φ_m is the magnetic flux through it.

QUESTIONS

1. Consider the four situations in Figure 27-1. Show in each case by use of Lenz's law that the direction of the current is as shown.
2. Consider the situation in Figure 27-1a, but suppose this time that the magnet is stationary and the loop moves toward it. Find the direction of current by use of Lenz's law.
3. Consider a region of space where there is a uniform magnetic induction parallel to the z-axis of a certain coordinate system. Does an observer who is moving along the z-axis detect any electric field? If so, what is its direction?
4. Repeat Question 3, but this time suppose that the observer is moving along: **(a)** the positive x-axis; **(b)** the positive y-axis.
5. Consider the situation in Figure 27-2. What is the direction of the current in loop B at the instant that the switch S is closed? What is its direction when the switch is subsequently opened? Use Lenz's law.
6. A steady current flows in a circular loop. Find the direction of the associated electric field at a point on the axis of the loop as seen by an observer who is traveling at a velocity **v**:
   (a) Directed along the axis of the loop.
   (b) Directed at right angles to the axis of the loop.
   (c) Directed along a line which makes an angle α with the axis.
7. Repeat Question 6, but suppose that the observer has in addition a certain *acceleration* along **v**.
8. A circular loop has a resistance R and is pulled through a *uniform* magnetic induction, which is directed perpendicularly to the plane of the loop. Will there be any current in the loop? Explain your answer in physical terms.
9. Is any force required to pull the loop of Question 8, assuming that it travels at a uniform velocity?
10. What is the direction of the induced emf in the lower segment of the rotating rectangular loop in Figure 27-20? What is it along the two vertical segments?
11. Consider, in Figure 27-22, a flat metal plate, which initially is at rest in an inhomogeneous magnetic induction **B**. Explain, in terms of induced currents, why it is necessary to exert a force in order to pull the plate to the right. (Such currents are known as *eddy currents.*)

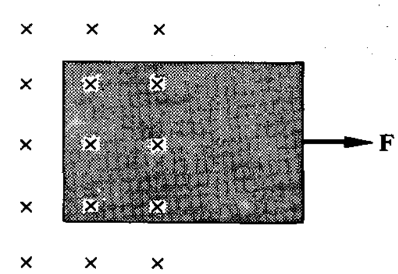


Figure 27-22

12. A small loop of wire is pulled at a velocity **v** between the pole pieces of a magnet as shown in Figure 27-23. What is the direction of the current in the loop just as it enters the region between the poles? What is the direction of this current on leaving this region?

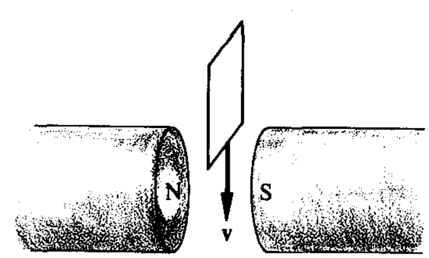


Figure 27-23

13. A particle of charge $q\,(>0)$ is traveling at a velocity **v** in the plane of a small conducting loop. At an instant when it is located as shown in Figure 27-24, why is the magnetic flux through the loop increasing? What is the direction of the current in the loop?

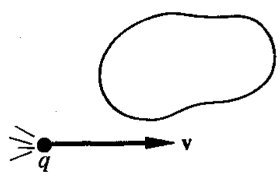


Figure 27-24

14. Consider the same situation as in Question 13, but this time as seen by an observer with respect to whom the particle is at rest so that the loop moves to the left at the velocity **v**. Why must there be a current in the loop this time also?
15. Explain in physical terms why, in light of Question 14, an emf can be induced in a conducting loop which moves through an *electrostatic field* for which

$$\oint \mathbf{E} \cdot d\mathbf{l} = 0 \qquad \text{(electrostatic field)}$$

so that apparently no emf can be induced.
16. A segment of wire is pulled through a uniform magnetic induction at a velocity **v**. Account for the emf developed between the ends of the wire in terms of the magnetic force on the charge carriers in the wire and their subsequent migrations.
17. Explain in physical terms the reason for the existence of an emf between the endpoints of a wire segment in a time-dependent magnetic induction $\mathbf{B}(t)$.
18. Suppose that the magnetic flux through a loop is increasing because a nonuniform **B**-field perpendicular to its plane is rising. Explain physically why this is the same as if the magnetic induction were constant (in time) and the loop had a velocity in a certain direction.
19. Suppose that there is a loop of wire of radius r along the circular path in Figure 27-16. Would the electrons in the wire still be accelerated to high speeds? Explain.
20. Describe what physical properties an electromagnet—such as that in Figure 27-16—must have so that it can accelerate electrons, as in a betatron, in a circle of fixed radius r.
21. If a particle of charge q moves through a time-independent magnetic induction, its energy does not change. Explain why the same is generally *not* true for a time-dependent **B**-field.
22. A particle of charge q is initially at rest in a time-dependent magnetic induction $\mathbf{B}(t)$. Explain why it will accelerate even though initially the magnetic force $q\mathbf{v} \times \mathbf{B}$ vanishes.

PROBLEMS

1. Consider a region of space in which there is a uniform **B**-field of strength 0.2 tesla. Assuming that this field is oriented along the positive x-axis of a certain Cartesian coordinate system, find the electric field as seen by an observer who is traveling at:
   (a) A velocity of 10 m/s along the z-axis.
   (b) A velocity of 100 m/s along the positive x-axis.
   (c) A velocity of 5.0 m/s along the negative y-axis.
2. A long, tightly wound solenoid has 200 turns per meter and carries a

current of 5.0 amperes. Calculate the electric field as seen by an observer who is traveling at a velocity of 50 m/s: **(a)** along the axis of the solenoid; **(b)** at right angles to its axis.

3. Two very long, straight wires are parallel, carry equal and opposite currents of 10 amperes, and are separated by a distance of 50 cm.
   (a) What is the magnetic induction at a point midway between them?
   (b) What is the electric field at a point midway between the wires as seen by an observer who is traveling at a velocity of 30 m/s parallel to the wires?
   (c) Repeat (b) for an observer who is traveling at a velocity of 20 m/s in the plane of the wires and perpendicular to them.

4. Two parallel wires separated by a distance d carry equal and opposite currents i. Calculate the electric field $\mathbf{E}$ at the point P, which is at a distance x from one of the wires, as seen by an observer with respect to whom the wires are traveling at the velocity $\mathbf{v}$ directed as shown in Figure 27-25.

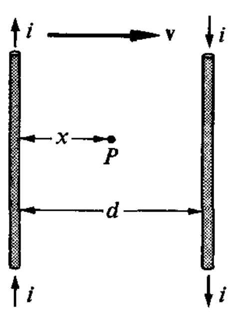


Figure 27-25

5. Repeat Problem 4, but suppose this time that the velocity $\mathbf{v}$ is parallel to the current on the left in Figure 27-25.

***6.** Suppose that $d\mathbf{E}$ represents the electrostatic field produced by an infinitesimal element of charge dq.
   (a) Show that the magnetic induction $d\mathbf{B}$ as seen by an observer with respect to whom this charge element has a velocity $\mathbf{v}$ is

$$d\mathbf{B} = \mu_0\epsilon_0\mathbf{v} \times d\mathbf{E}$$

   (b) Making use of this result, show by integration that if $\mathbf{E}$ is the electric field produced by a certain distribution of charge, then the magnetic induction $\mathbf{B}$ as seen by an observer with respect to whom the charge has a velocity $\mathbf{v}$ is

$$\mathbf{B} = \epsilon_0\mu_0\mathbf{v} \times \mathbf{E}$$

   In Chapter 29, it will be seen that $\epsilon_0\mu_0 = 1/c^2$, where c is the velocity of light. Hence $B \approx vE/c^2$ and is generally very small compared to E/c for speeds $v \ll c$.

7. Consider a sphere of radius a, which carries a uniform charge of density ρ_0. If the sphere is traveling at a velocity $\mathbf{v}$ find, by use of the results of Problem 6, the magnetic induction $\mathbf{B}$ inside and outside the sphere.

8. A circular coil of wire of radius a has N turns and carries a current i. Calculate the electric field $\mathbf{E}$ at a point P on the axis of the coil as seen by an observer with respect to whom the coil has a velocity $\mathbf{v}$ parallel to a diameter. Assume P is at a distance x from the plane of the loop on the side where $\mathbf{B}$ is directed toward the loop.

9. In Figure 27-7a, suppose that $B = 0.1$ tesla, $v = 10$ m/s, and $l = 20$ cm. **(a)** What is the electric field inside the rod? **(b)** What emf is developed between its ends? Which end acquires the positive charge?

10. In Figure 27-8a suppose that $v = 10$ m/s, $l = 10$ cm, $B = 0.5$ tesla, and

$R = 10\,\Omega$. Calculate (a) the emf in the circuit; (b) the current in the circuit; and (c) the power dissipated.

11. A runner whose height is 2.0 meters runs east at a speed of 20 km/hr perpendicular to the earth's **B**-field. Assuming that the horizontal component of **B** has a magnitude 5.0×10^{-5} tesla, what emf is developed between the top of his head and the soles of his feet?

12. Repeat Problem 11, but assume this time a "being" on a *neutron star* where the **B**-field has a strength of 10^9 tesla. Assume the same values for "its" height and speed as above.

13. If, for the physical situation in Figure 27-8a, $B = 0.5$ tesla, $l = 10$ cm, $v = 5$ m/s, and a current of 15 mA flows, calculate:

(a) The resistance in the circuit.

(b) The power dissipated.

(c) The force required to keep the rod moving.

(d) The power expended by the external agent and compare with your results to (b). Account for any differences.

14. If, for the physical situation in Figure 27-8a, $B = 0.1$ tesla, $l = 5$ cm, and $R = 2.0\,\Omega$, and a steady current of 2.0 mA flows, calculate (a) the velocity of the rod and (b) the rate at which the area of the loop increases.

15. A conducting rod of length l and resistance R slides down the conducting rails on the edges of an inclined plane of angle α as shown in Figure 27-26. Assume that a uniform magnetic induction **B** is directed vertically downward as shown in the figure, that the total electrical resistance is R, and that no frictional forces act. At an instant when the rod has the velocity **v**, as shown:

(a) Show that the emf developed around the conducting path is $Blv \cos \alpha$.

(b) Calculate the magnitude and direction of the current.

16. Consider again the situation in Figure 27-26 and suppose that the rod has a mass m.

(a) Show that the magnetic force F on the rod is

$$F = \frac{B^2 l^2 v}{R} \cos \alpha$$

and find the direction of this force.

(b) Show that the equation of motion for the rod is

$$mg \sin \alpha - \frac{B^2 l^2 v \cos^2 \alpha}{R} = m\frac{dv}{dt}$$

(c) Find the terminal velocity.

17. A conducting rod of length l is pivoted at a point P and rotates with uniform angular velocity ω in a plane at right angles to a uniform induction **B** (see Figure 27-27).

(a) Show that the emf $d\mathscr{E}$ developed across an infinitesimal

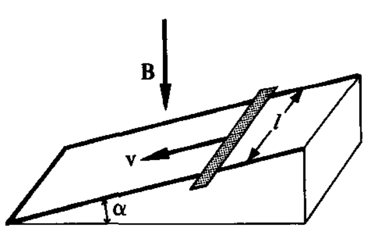


Figure 27-26

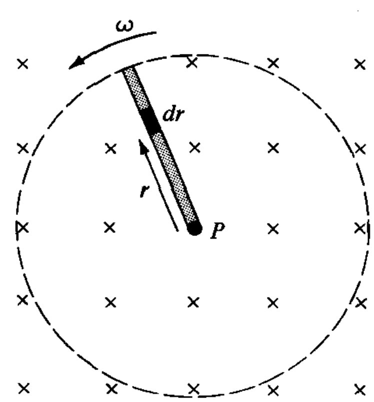


Figure 27-27

element of the rod of length dr at a distance r from P is

$$d\mathscr{E} = -\omega Br\,dr$$

Discuss the meaning of the minus sign in terms of the direction of the emf development between the ends of the rod.

(b) By integration, calculate the magnitude and the direction of the total emf developed between the ends of the rod.

18. A conducting rod of length l is pivoted at a point P and the other end slides along a circular conducting track with a uniform speed v. Assume that there is a uniform magnetic induction $\mathbf{B}$ directed as shown in Figure 27-28, that P and the track are connected by a resistor R, and that all other electrical resistance in the circuit is negligible.

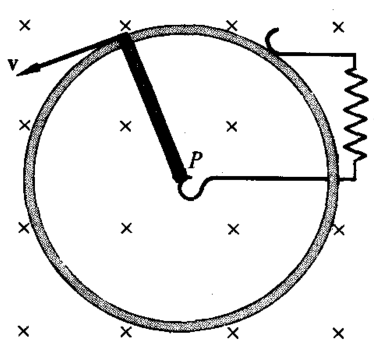


Figure 27-28

(a) Calculate the magnitude and the direction of the emf developed in the circuit.

(b) What is the current through the resistor?

(c) How much power is dissipated in the resistor?

(d) What torque is required to keep the rod rotating?

19. A circular, conducting disk of radius a rotates with uniform angular velocity ω about its axis. Assuming the existence of a uniform magnetic induction $\mathbf{B}$ parallel to the axis, show that the emf developed between the center of the disk and its rim is

$$\mathscr{E} = \frac{\omega a^2 B}{2}$$

(*Hint*: Think of the disk as a collection of rods, such as the one in Problem 17.) (*Note*: This device used for generating an emf was first studied by Faraday and is known as the *Faraday disk dynamo*.)

20. Consider again the physical situation described in Problem 19. Calculate the electric field at each point of the disk due to the charge separation that occurs. Assume the sense of rotation in Figure 27-28.

21. A conducting rod of length l is located near a long, straight wire, which carries a current i, as in Figure 27-29. If the rod is pulled at a velocity $\mathbf{v}$ along the direction of the current, calculate in terms of a, l, α, v, and i the emf developed between the ends of the rod.

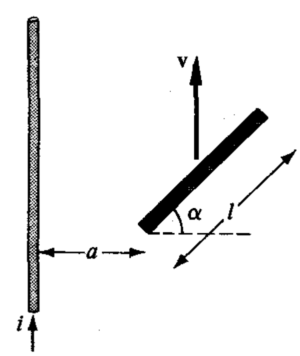


Figure 27-29

22. Consider again the situation in Figure 27-29, but suppose this time that the rod is pulled to the right with a velocity $\mathbf{v}$. Assuming that at $t = 0$ the lower end of the rod is at the distance a from the current, calculate, as a function of time, the emf developed between the ends of the rod.

23. Suppose that the rotating current

loop in Figure 27-20 is a square of side l. At the instant depicted in the figure:

(a) What emf is developed along the bottom segment of the loop?

(b) What is the magnitude and direction of the emf developed in each of the vertical segments of the loop?

(c) Make use of (a) and (b) to calculate the total emf developed around the loop and compare your result with that in (27-25).

24. Consider, in Figure 27-30, a square loop of wire of length l and resistance R, which is forced to travel at the velocity $\mathbf{v}$ directed as shown. Assuming that at $t = 0$ it first enters the region of a uniform magnetic induction $\mathbf{B}$, and that $a > l$:

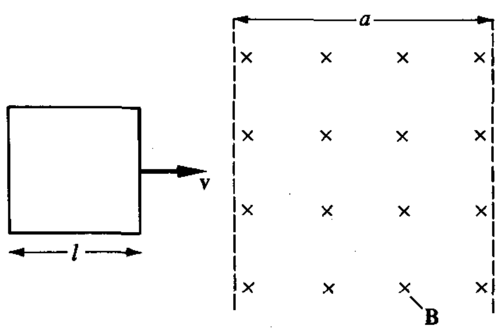


Figure 27-30

(a) Calculate the strength and the direction of the current in the loop as a function of time for the time intervals (1) $t \leq l/v$; (2) $l/v \leq t \leq a/v$; and (3) $t > a/v$.

(b) Calculate the force required to keep the loop moving at the constant speed v during each of the above time intervals.

25. Show by use of (27-12) that the emf developed around a rigid but closed conducting loop which moves at a velocity $\mathbf{v}$ in a uniform time-independent magnetic induction $\mathbf{B}$ vanishes. *Hint*: For this situation, (27-12) may be written as

$$\mathscr{E} = (\mathbf{v} \times \mathbf{B}) \cdot \oint d\mathbf{l}$$

***26.** Show explicitly that in the presence of a *time-dependent* **B**-field, the work required to take a charged particle between any two points depends *in general* on the path connecting them. (*Hint*: Why is the work required to take a particle around a closed path not zero except in electrostatics?)

27. A conducting loop of radius 10 cm is in a uniform magnetic induction, which is directed as shown in Figure 27-31 and which increases at the constant rate of 0.02 T/s.

(a) What is the direction of the emf developed around the loop?

(b) What is the magnitude of the emf around the loop?

(c) If the resistance of the loop is 100 Ω, what current flows?

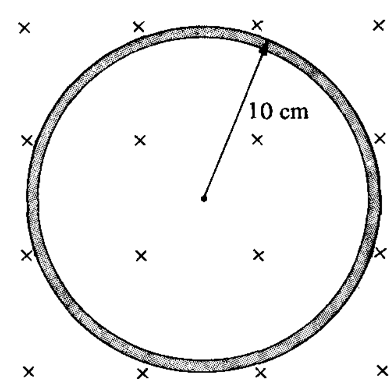


Figure 27-31

28. Consider again the situation in Figure 27-31, but suppose that this time the magnetic induction decreases at the rate of 0.2 T/s. Calculate the magnitude and the direction of the current in the loop.

29. A very long ideal solenoid of radius a has n turns per unit length and carries a current $i = i_0 + \alpha t$, where i_0 and α are positive constants.

(a) Calculate the magnetic flux

$\Phi_m(t)$ through a coaxial conducting loop of radius b ($>a$). See Figure 27-32.

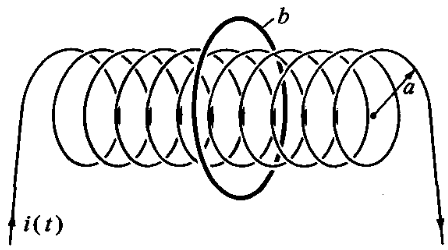


Figure 27-32

(b) Calculate the emf developed around the loop.
(c) If the loop has a resistance R, find the magnitude and the direction of the current in the loop.

30. A rectangular loop of wire of sides a and b has a resistance R and is subjected to a uniform magnetic induction perpendicular to its plane and varying in time in accordance with the formula

$$B = B_0 \cos \omega t$$

where B_0 and ω are constants. Calculate the current in the loop as a function of time.

31. Suppose that the current i in the long wire in Figure 27-19 varies as $i = i_0 - \alpha t$, with i_0 and α positive constants.
(a) Calculate the magnitude and direction of the emf developed in the rectangular loop.
(b) If the loop is cut at some point, what is the potential difference between the cut ends?

*32. A circular loop of wire of radius a is in a uniform magnetic induction **B** and is attached to a resistor R as shown in Figure 27-33.
(a) What is the magnetic flux through the circular loop?
(b) Suppose that the loop is quickly rotated in a short time interval Δt by 180° about the axis AC

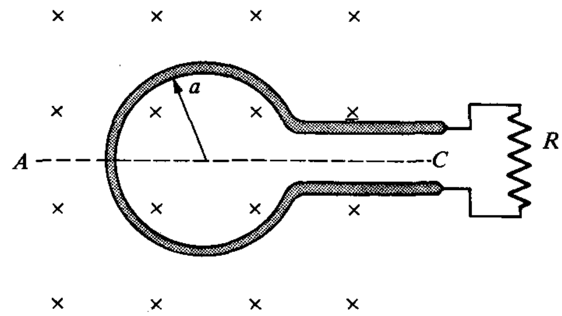


Figure 27-33

while the rest of the circuit maintains its original position. What is the direction of the current in the resistor R?
(c) Show that the total charge Q that flows through the resistor is

$$Q = \frac{2\pi a^2 B}{R}$$

*33. Suppose, in Figure 27-19, that the conducting loop is circular of radius a and with its center at a perpendicular distance b ($>a$) from the long wire.
(a) What is the magnetic flux through the loop?
(b) Calculate the emf developed around the loop.

*34. Show by use of Faraday's law that the electric field cannot drop abruptly to zero, in a direction perpendicular to itself. (*Hint*: Assume that, on the contrary, the electric field does drop abruptly to zero, evaluate Faraday's law for the rectangular path *abcd* in Figure 27-34, and consider the limit as the sides *ad* and *bc* become vanishingly small.)

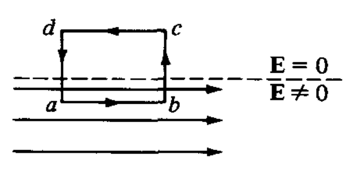


Figure 27-34

*35. Show by use of Faraday's law that even if magnetic monopoles

existed, so that the magnetic flux out of a closed surface did not vanish, the relation

$$\frac{d}{dt}\oint_S \mathbf{B}\cdot d\mathbf{S} = 0$$

would still be valid.

***36.** A circular loop of wire is in a time-dependent magnetic induction $\mathbf{B}(t)$ directed perpendicular to its plane. By starting with Faraday's law show that the direction of the emf developed in the coil is the same as that obtained by use of Lenz's law.

37. Consider, in Figure 27-35, a cylindrical region of radius a in which there is a uniform magnetic induction $\mathbf{B}(t)$, with $B(t) = B_0 - \alpha t$, with B_0 and α positive constants.

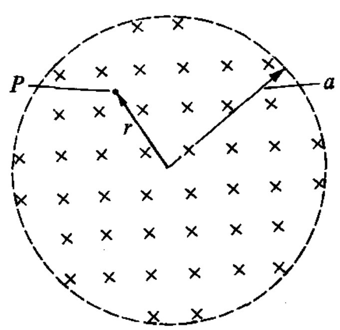


Figure 27-35

(a) What is the instantaneous force on a proton of charge q when it is at rest at the point P?

(b) What is the instantaneous acceleration of an electron if it is placed at rest at this point? Assume that $r = 50$ cm and $\alpha = 0.02$ T/s.

38. Consider again the cylindrically symmetric magnetic induction $B(t) = B_0 - \alpha t$ in Figure 27-35. Suppose that a loop of wire of resistance R and radius $a/2$ is placed so that it lies in the plane of the figure and is totally immersed in the uniform magnetic induction.

(a) Calculate the magnitude and the direction of the current in the loop.

(b) Suppose that the loop is replaced by one that is a square of side $a/2$. What changes, if any, are required for your answer to (a)?

***39.** Show that the electric field $\mathbf{E}$ at any point inside the dotted circle in Figure 27-35 is

$$\mathbf{E} = \frac{1}{2}\mathbf{r}\times\frac{d\mathbf{B}}{dt}$$

where $\mathbf{B}$ is the spatially uniform magnetic induction and $\mathbf{r}$ is a vector from the center to the point under consideration. What is the equation of motion for a particle of charge q and mass m in this region?

***40.** For the situation in Figure 27-18, suppose that the external field $\mathbf{B}(t)$ is not constant but varies in time and that the total resistance in the conducting path is R.

(a) Show that the emf developed around the closed conducting path is

$$\mathscr{E} = B(t)lv + xl\frac{dB}{dt}$$

(b) For the special case $B(t) = B_0 - \alpha t$, where B_0 and α are positive constants, calculate the force F required to keep the wire moving at the velocity v. Assume that at $t = 0$, $x = x_0$.

(c) Compare the power Fv expended by the external agent with that dissipated in the resistor and account for the difference in physical terms.

28 Inductance and magnetic materials

First we must inquire whether the elements are eternal or subject to generation and destruction; for when this question has been answered their number and character will be manifest.

ARISTOTLE

28-1 Introduction

In our study of dielectrics it was convenient to introduce first the notion of capacitance and then to use capacitors as a tool to measure the electric properties of insulators. Our study of magnetic materials will proceed in a similar way. This time we first introduce the concept of *inductance* and then show how the magnetic properties of matter can be measured by use of *inductors.*

28-2 Self-inductance

If a time-dependent current is generated in one of two neighboring conducting loops, then because of the changing magnetic flux, a current will, in general, also be induced in the other. Similarly, if a time-dependent current is generated in an isolated loop, the changing magnetic flux produces an additional emf around the loop itself. It is convenient to characterize this

induced emf in a loop—whether it be due to its own changing current or to that in a neighboring one—in terms of certain quantities known as the *coefficients of inductance.* The purpose of this and Section 28-3 is to define these quantities.

Consider, in Figure 28-1, a loop of wire around which flows a time-dependent current $i \equiv i(t)$. Associated with this current, there will be a magnetic induction $\mathbf{B}(t)$, which in turn produces a certain magnetic flux Φ_m through the loop. According to (25-5), $\mathbf{B}(t)$ is directly proportional to the current. Hence we may write

$$\Phi_m = Li \tag{28-1}$$

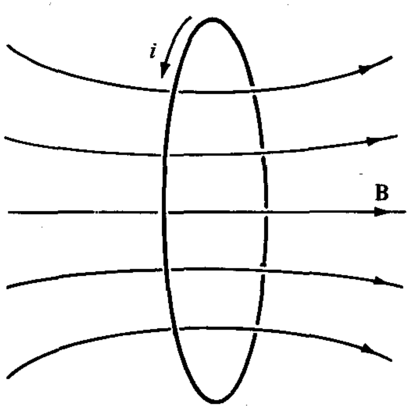


Figure 28-1

with L a certain proportionality constant, which is known as the *inductance* or the *self-inductance* of the loop. Note that L is independent of i; it depends only on the shape and size of the loop. Following convention we shall always take the positive sense of the normal to the loop to be related to the direction of the current by the right-hand rule. With this choice, the magnetic flux Φ_m in (28-1) will be positive, as will be the self-inductance L of the loop.

For the special case of a closely packed coil or an ideal solenoid for which the magnetic flux Φ_m through *each* of the N turns is the same, the total flux through the coil is $N\Phi_m$. For this case, (28-1) assumes the form

$$N\Phi_m = Li \tag{28-2}$$

Note that in both (28-1) and (28-2) Φ_m is the flux through a *single* turn of the coil.

By use of the above definition of inductance, the emf across a coil of wire, or an *inductor,* as it is known, can be expressed directly in terms of L. Assuming that a current i flows around the loop, we find, by substituting (28-1) into Faraday's law, that

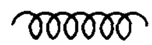
$$\mathscr{E} = -\frac{d\Phi_m}{dt} = -\frac{d}{dt}(Li)$$

and, since L is a time-independent constant, this may be expressed equivalently as

$$\mathscr{E} = -L\frac{di}{dt} \tag{28-3}$$

In other words, the flow of a current $i(t)$ around a loop produces the same effect as if a "battery" of emf $(-L\,di/dt)$ had been introduced into the loop. Although (28-3) has been derived for a single loop, it is easy to confirm by use of (28-2) that it is equally valid for a closely packed coil or an ideal solenoid.

When used in electric circuits, the symbol



will always be used to represent an inductor. Circuits containing inductors will be discussed in Section 28-4.

According to the definition in (28-1), the unit of inductance is the $\text{T-m}^2/\text{A} = \text{Wb/A}$. It is convenient to define a unit of inductance called the *henry* (abbreviated H) by the relation

$$1\,\text{H} = 1\,\text{Wb/A}$$

The related units of the *millihenry* $\text{mH} = 10^{-3}\,\text{H}$ and the *microhenry* $\mu\text{H} = 10^{-6}\,\text{H}$ are also frequently used. This unit is named in honor of Joseph Henry (1797–1878), a contemporary of Faraday, who independently carried out studies of inductive effects associated with time-dependent currents.

Example 28-1 A very long, ideal solenoid has N turns and a length l. Calculate its inductance assuming that each turn is circular and has a radius a.

Solution Assuming that a current i flows, the magnetic induction $\mathbf{B}$ is uniform and parallel to the axis of the coil. According to (25-9), its magnitude is

$$B = \frac{\mu_0 Ni}{l}$$

Hence, the magnetic flux Φ_m through a single turn of the coil is

$$\Phi_m = B\pi a^2 = \frac{\mu_0 iN\pi a^2}{l}$$

and substitution into (28-2) leads to

$$L = \frac{N\Phi_m}{i} = \frac{N\mu_0 iN\pi a^2}{li} = \frac{\mu_0 N^2\pi a^2}{l} \tag{28-4}$$

For example, if $a = 1\,\text{cm}$, $N = 10^3$, and $l = 20\,\text{cm}$, then

$$L = \frac{\mu_0 N^2\pi a^2}{l} = \frac{(4\pi\times10^{-7}\,\text{T-m/A})\times(10^3)^2\times\pi\times(10^{-2}\,\text{m})^2}{0.2\,\text{m}} = 2.0\times10^{-3}\,\text{H}$$

Example 28-2 A toroid of inner radius a and outer radius b has N turns and a rectangular cross section of width c. Assuming that a current i flows, calculate:
(a) The magnetic flux through a single turn of the toroid.
(b) The self-inductance of the toroid.

Solution Figure 28-2a shows a top view of the toroid and Figure 28-2b a cross-sectional view without the wire.

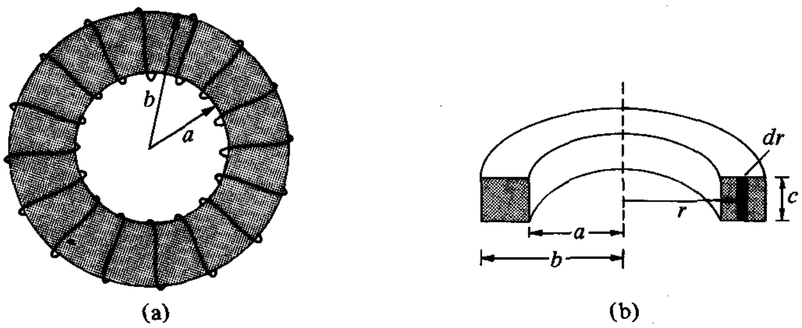


Figure 28-2

(a) According to the argument in Chapter 25, the lines of magnetic induction in a toroid are circles concentric with its axis. The magnitude of the **B**-field at a distance r $(a < r < b)$ from its center is given by (25-15); that is,

$$B = \frac{\mu_0 N i}{2\pi r}$$

Hence the magnetic flux $d\Phi_m$ through a small rectangular area of width dr and height c, and at a distance r from the axis is

$$d\Phi_m = Bc\,dr = \frac{\mu_0 N i}{2\pi r}c\,dr$$

The flux Φ_m through a single turn is then obtained by integration

$$\Phi_m = \int d\Phi_m = \int_a^b \frac{\mu_0 N i}{2\pi r}c\,dr = \frac{\mu_0 N i c}{2\pi}\int_a^b \frac{dr}{r} = \frac{\mu_0 N i c}{2\pi}\ln r\Big|_a^b = \frac{\mu_0 N i c}{2\pi}\ln\frac{b}{a}$$

since $d[\ln r]/dr = 1/r$.
(b) Substituting the above formula for Φ_m into (28-2) we obtain

$$L = \frac{N\Phi_m}{i} = \frac{\mu_0 N^2 c}{2\pi}\ln\frac{b}{a} \qquad \textbf{(28-5)}$$

28-3 Mutual inductance

A time-dependent current in a coil will, in general, produce a changing magnetic flux not only through itself but also through any other nearby coil. To study this possibility consider, in Figure 28-3, two *closely packed* coils, which carry the currents i_1 and i_2 and consist of N_1 and N_2 turns of wire, respectively. Let Φ_{12} be the magnetic flux through a single turn of coil 1 due

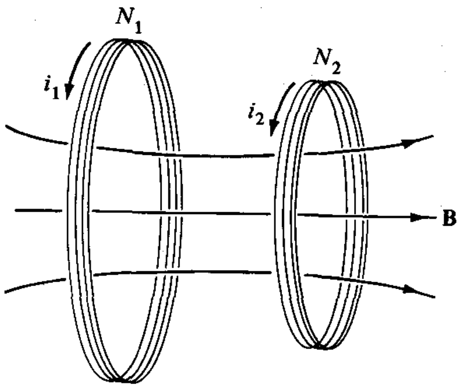


Figure 28-3

to the current i_2 in coil 2 and Φ_{21} the corresponding flux through a single turn of coil 2 due to i_1. It is important to note that Φ_{12}, for example, is *not* the total magnetic flux through a single turn of coil 1, since its own current i_1 produces a flux as well. The total magnetic flux through coil 1 due to i_2 is $N_1\Phi_{12}$, and, correspondingly, $N_2\Phi_{21}$ represents the total magnetic flux through coil 2 due to i_1.

Now as in the discussion of self-inductance, the magnetic fluxes $N_1\Phi_{12}$ and $N_2\Phi_{21}$ through the two coils will be proportional, respectively, to i_2 and i_1. Accordingly, we define the coefficients of mutual inductance M_{12} and M_{21} by

$$\begin{aligned} M_{21}i_1 &= N_2\Phi_{21} \\ M_{12}i_2 &= N_1\Phi_{12} \end{aligned} \tag{28-6}$$

so that M_{12} and M_{21} are current-independent proportionality constants; they depend only on the geometric parameters characterizing the two loops and their relative separation and orientation. By contrast to the coefficient of self-inductance L, which characterizes the flux through a loop due to its own current, the mutual inductance M_{12} characterizes the flux through coil number 1 due to a current in a neighboring coil 2. Correspondingly, M_{21} represents the flux through coil 2 due to a unit current in coil 1.

We shall always assume in the following that each coil has been assigned a sense of direction such that a positive current in one produces a positive flux through the other. If, for example, the current i_1 in Figure 28-3 is positive, then if the sense of the other loop is along i_2, the flux Φ_{21} through the latter is positive. With this choice, the sign of Φ_{21} is always the same as i_1, and similarly for Φ_{12} and i_2. It follows then by reference to (28-6) that the coefficients of mutual inductance M_{12} and M_{21} are invariably positive.

Detailed calculations of the coefficients of mutual inductance for a variety of coils under a variety of relative positions show that the the two coefficients M_{12} and M_{21} are always the same. Hence we shall write

$$M = M_{12} = M_{21} \tag{28-7}$$

and the symbol M will be used to represent the mutual inductance of either coil. The relation in (28-7) is extremely useful in practice, and is a special case of a more general relation, which is known as a *reciprocity relation.* We emphasize that the validity of (28-7) is *not* dependent on the currents or the fluxes through the two coils. In general, the fluxes Φ_{12} and Φ_{21} will *not* be equal to each other despite (28-7).

The emf $\mathscr{E}_1$ developed in coil 1 due to the changing current in coil 2 may be calculated by use of Faraday's law. Since $N_1\Phi_{12}$ is the total flux through coil 1 due to the current i_2 in coil 2, it follows that

$$\begin{aligned}\mathscr{E}_1 &= -\frac{d}{dt}(N_1\Phi_{12}) = -\frac{d}{dt}(M_{12}i_2)\\ &= -M\frac{di_2}{dt}\end{aligned} \tag{28-8}$$

where in the last equality we have used (28-7) and the fact that M is a time-independent constant. In a similar way, the emf $\mathscr{E}_2$ developed in coil 2 due to i_1 is

$$\mathscr{E}_2 = -M\frac{di_1}{dt} \tag{28-9}$$

Example 28-3 A large, closely packed coil of N_2 turns completely surrounds a very long, ideal solenoid of length l, radius a, and N_1 turns, as in Figure 28-4. Calculate the mutual inductance M between the coils.

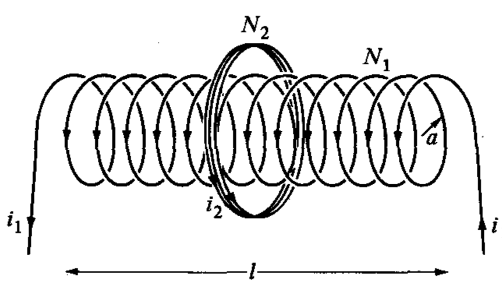


Figure 28-4

Solution Assuming a current i_1 in the solenoid, the magnetic induction **B** inside is uniform, parallel to its axis, and has the strength $B = \mu_0 N_1 i_1/l$. Outside the solenoid the magnetic induction vanishes. The magnetic flux Φ_{21} through a single turn of the large coil is thus

$$\Phi_{21} = B\pi a^2 = \left(\frac{\mu_0 N_1 i_1}{l}\right)\pi a^2$$

and hence, by use of (28-6) and (28-7), the mutual inductance M is found to be

$$M = \frac{N_2\Phi_{21}}{i_1} = \frac{N_2}{i_1}\left(\frac{\mu_0 N_1 i_1}{l}\right)\pi a^2 = \frac{\mu_0 \pi a^2}{l} N_1 N_2$$

It is interesting to note that even though it is very difficult to calculate the flux Φ_{12} through the solenoid directly, by use of this formula for M and (28-7) the calculation

is straightforward. For, according to (28-6),

$$\Phi_{12} = \frac{M_{12}i_2}{N_1} = \frac{Mi_2}{N_1} = \frac{\mu_0\pi a^2 N_2}{l}\, i_2$$

where the final equality follows by use of the above formula for M. By contrast, the calculation of Φ_{12} by direct integration is extraordinarily complex!

Example 28-4 A rectangular loop of wire of dimensions l and $(d-2a)$ lies in the plane of two very long, parallel wires, which comprise part of a circuit (see Figure 28-5). Calculate the mutual inductance between the loops.

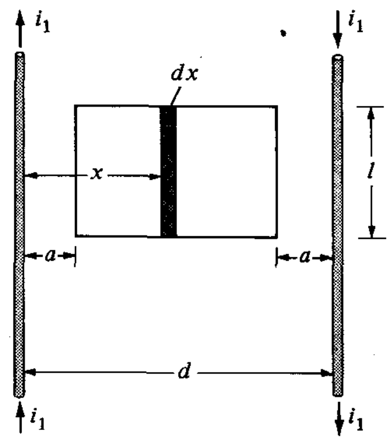


Figure 28-5

Solution Assuming that a current i_1 flows in the long wires, it follows that the flux $d\Phi_{21}$ through the small element of area of length l and thickness dx of the loop is

$$\begin{aligned} d\Phi_{21} &= Bl\,dx \\ &= \frac{\mu_0 i_1}{2\pi}\left(\frac{1}{x} + \frac{1}{d-x}\right) l\,dx \end{aligned}$$

The total flux Φ_{21} through the small loop is found, by integration from $x = a$ to $x = d - a$, to be

$$\begin{aligned} \Phi_{21} &= \int d\Phi_{21} = \frac{\mu_0 i_1 l}{2\pi}\int_a^{d-a} dx\left[\frac{1}{x} + \frac{1}{d-x}\right] \\ &= \frac{\mu_0 i_1 l}{2\pi}[\ln x - \ln(d-x)]\Big|_a^{d-a} \\ &= \frac{\mu_0 i_1 l}{\pi}\ln\left(\frac{d-a}{a}\right) \end{aligned}$$

Finally, using the value $N_2 = 1$ and (28-6), we obtain

$$M = \frac{\Phi_{21}}{i_1} = \frac{\mu_0 l}{\pi}\ln\left(\frac{d-a}{a}\right)$$

28-4 The *R-L* circuit

Consider, in Figure 28-6, an inductor of inductance L in series with a resistor R, both connected across a battery of emf $\mathscr{E}$. Suppose that the circuit also contains a switch S, which is closed at the initial instant $t = 0$. According to Lenz's law, the inductor strives to oppose any change of magnetic flux through itself. Therefore, just after the switch is closed, an emf will be produced in the inductor and its sense will be such as to oppose, at least momentarily, the flow of electric current in the circuit. Initially, therefore, the current i through the battery will vanish. After a long time, however, the current becomes steady and then the inductor no longer influences the flow of current. This is to be contrasted with the R-C circuit analyzed in Section 23-6, where initially the current assumes its maximum value and only after the capacitor is fully charged does the flow of current cease.

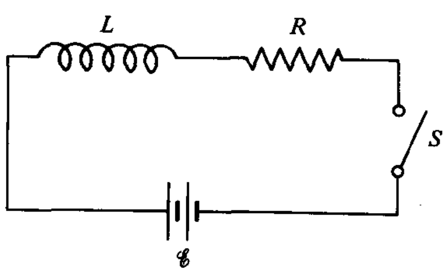


Figure 28-6

According to (28-3), an inductor through which flows a current i behaves as if it were a battery of emf $(-L\,di/dt)$. Hence the circuit equation for the R-L circuit is

$$Ri = \mathscr{E} - L\frac{di}{dt}$$

and this can be expressed equivalently as

$$L\frac{di}{dt} + Ri = \mathscr{E} \qquad \textbf{(28-10)}$$

This is the basic equation for the R-L circuit. Its solution, subject to the initial condition $i = 0$ at $t = 0$, constitutes a complete description for the current at any time t.

To solve (28-10), let us first express it in the form

$$\frac{di}{i - \mathscr{E}/R} = -\frac{dt\,R}{L}$$

and then integrate:

$$\ln\left(\frac{\mathscr{E}}{R} - i\right) = -\frac{tR}{L} + \alpha$$

In order to satisfy the initial condition $i = 0$ at $t = 0$, the integration constant α must be $\ln(\mathscr{E}/R)$. Hence the final formula for the current in the *R-L* circuit is

$$i = \frac{\mathscr{E}}{R}(1 - e^{-Rt/L}) \tag{28-11}$$

Figure 28-7 shows a plot of i as a function of time. The current rises steadily from its initial value of zero and approaches its saturation value $\mathscr{E}/R$ asymptotically. In other words, after a long time, the current in the circuit assumes the value which it would have had if the inductance were *not* in the circuit. In this context by a "long time" is meant a period of time that is long compared to the *R-L* "time constant" τ_L, defined by

$$\tau_L = \frac{L}{R} \tag{28-12}$$

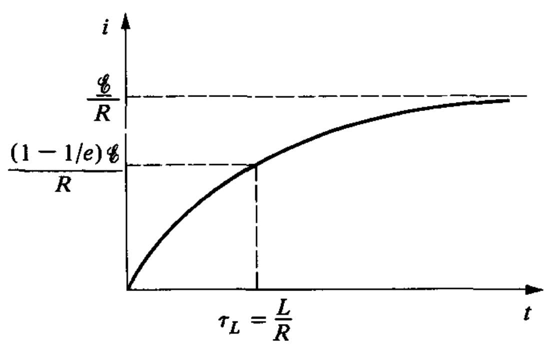


Figure 28-7

As shown in the figure τ_L represents the time interval from the closing of the switch to the instant at which the current has risen to $(1 - 1/e)$ of its final value, $\mathscr{E}/R$.

It is interesting to contrast the behavior of this *R-L* circuit with the corresponding behavior of the *R-C* circuit. In the latter the capacitor acts initially as a short, so that at first the current flows as if the capacitor were not present. After a long time, when the charge on the capacitor has built up, it acts as a large resistor and prevents the flow of current in the circuit. By contrast, in the *R-L* circuit, the inductor opposes any changes in magnetic flux through its coils so that initially it acts as a very large resistor, and stops any current flow. After a long time, however, the current achieves the steady value it would have if the inductance were not present, so that, effectively, it acts as a short.

Example 28-5 A 20-volt battery is suddenly connected across a 0.2-henry inductor that has a resistance of 100 Ω. Calculate:

(a) The *R-L* time constant.
(b) The potential drop across the inductor at $t = \tau_L$.
(c) The final current.

Solution

(a) Substituting the given values for R and L into (28-12) we obtain

$$\tau_L = \frac{L}{R} = \frac{0.2\text{ H}}{100\,\Omega} = 2.0 \times 10^{-3}\text{ s}$$

(b) The potential drop across the inductor is $(\mathscr{E} - Ri)$. Making use of (28-11) we obtain then, at $t = \tau_L$,

$$\mathscr{E} - Ri = \mathscr{E}e^{-t/\tau_L} = \frac{\mathscr{E}}{e} = 0.37 \times 20\text{ V} = 7.4\text{ V}$$

since $1/e \cong 0.37$ and the emf of the battery is 20 volts.

(c) After a long time, the current $i(\infty)$ in the circuit is the same as if the inductor were not present. Thus

$$i(\infty) = \frac{\mathscr{E}}{R} = \frac{20\text{ V}}{100\,\Omega} = 0.2\text{ A}$$

Example 28-6 Consider the two-loop network in Figure 28-8. Assuming that at $t = 0$ the switch S is closed, calculate:

(a) The initial current through the battery.
(b) The initial potential drop across the inductor.
(c) The final currents through R_1 and R_2.

Solution

(a) Just after the switch is closed, the inductor will not allow any current to pass and thus acts as an infinite resistor. Initially, then, the current will flow along the path $ABCDEFA$, and thus according to Ohm's law the current i_0 through the battery is

$$i_0 = \frac{\mathscr{E}}{R_1 + R_2}$$

(b) The potential drop V_L across the inductor must be the same as that across R_2. Making use of the result of (a), we find that

$$V_L(0) = R_2 i_0 = \frac{R_2\mathscr{E}}{R_1 + R_2}$$

(c) After a long time, steady-state conditions prevail, and the potential drop across the inductor, $L\,di/dt$, will vanish. Thus, there will be no current through R_2 since the potential drop across R_2 must be the same as that across L. The final value for the current $i(\infty)$ through R_1 is therefore

$$i(\infty) = \frac{\mathscr{E}}{R_1}$$

Example 28-7 Consider the circuit in Figure 28-9 and suppose that at $t = 0$ the switch S is closed. Calculate:

(a) The initial current through the battery.
(b) The final current through the battery.
(c) The final charge Q_f on the capacitor.

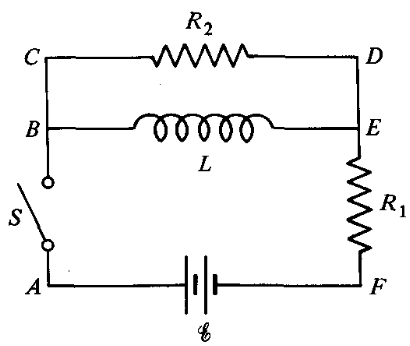


Figure 28-8

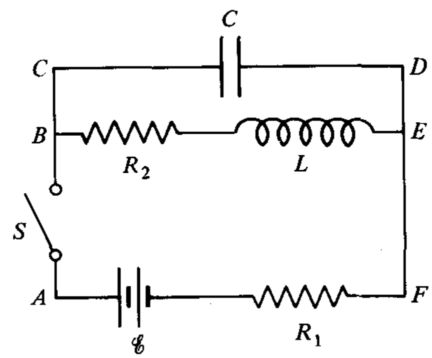


Figure 28-9

Solution

(a) Just as the switch is closed, the capacitor acts as a short, and the inductor behaves as if it were a resistor of infinite resistance. Hence, the initial current $i(0)$ travels along the path $ABCDEFA$, and since the total resistance in this path is R_1 we conclude that

$$i(0) = \frac{\mathscr{E}}{R_1}$$

(b) After a long time, when steady-state conditions prevail, the capacitor acts as an infinite resistor and the inductor acts as a short. Accordingly, the final current $i(\infty)$ follows the path $ABEFA$ and, since the total resistance along this path is $(R_1 + R_2)$, it has the value

$$i(\infty) = \frac{\mathscr{E}}{R_1 + R_2}$$

(c) Since in the final state the potential drop across the inductor vanishes, the potential across the capacitor must be the same as that across R_2. Hence

$$\frac{Q_f}{C} = R_2 i(\infty)$$

or, in other words, the final charge Q_f on the capacitor is

$$Q_f = CR_2 i(\infty) = \frac{CR_2\mathscr{E}}{R_1 + R_2}$$

where the final equality follows by use of the result of (b).

28-5 Magnetic energy

In Chapter 22 we found that associated with a charged capacitor there is the electrical energy $U_E = Q^2/2C$. The purpose of this section is to establish the analogous formula for inductors:

$$U_M = \frac{1}{2} Li^2 \tag{28-13}$$

where U_M is the magnetic energy stored in an inductor L through which flows a current i.

By analogy to the method used to derive (23-25), let us multiply the circuit equation (28-10) for the circuit in Figure 28-6 by the current i:

$$\mathscr{E}i = Ri^2 + iL\frac{di}{dt}$$

Since the inductance L is constant, and $d(i^2)/dt = 2i\,di/dt$, this may be written equivalently as

$$\mathscr{E}i = Ri^2 + \frac{d}{dt}\left(\frac{1}{2}Li^2\right)$$

Now the term $\mathscr{E}i$ represents the rate at which the battery carries out work on the other circuit elements. Furthermore, as we saw in the analogous derivation involving the R-C circuit, the term Ri^2 is the rate at which heat is dissipated in the resistor. It follows then from the ideas of energy conservation that the term $d(Li^2/2)/dt$ must represent the rate at which *energy is being stored in the magnetic field of the inductor.* The validity of (28-13) is thus established.

To illustrate, let us consider the special case of a long ideal solenoid of length l, radius a, and N turns. According to (28-4), the self-inductance L of the coil is $\mu_0 N^2\pi a^2/l$. Hence, assuming a current i flows, we find that

$$U_M = \frac{1}{2}Li^2 = \frac{1}{2}\frac{\mu_0 N^2\pi a^2}{l}i^2 \tag{28-14}$$

Example 28-8 Consider the R-L circuit in Figure 28-6 for the special parameter values $\mathscr{E} = 100$ volts, $R = 50\ \Omega$, and $L = 0.1$ henry. Calculate the maximum value for the energy stored in the inductor.

Solution Since the current in the circuit is a monotonically increasing function (see Figure 28-7), it follows by use of (28-13) that the maximum value for U_M occurs after a long time. Since the saturation value $i(\infty)$ for the current is $\mathscr{E}/R$, we obtain

$$U_M = \frac{1}{2}L[i(\infty)]^2 = \frac{1}{2}L\left(\frac{\mathscr{E}}{R}\right)^2 = \frac{1}{2}\times(0.1\text{ H})\times\left(\frac{100\text{ V}}{50\Omega}\right)^2 = 0.2\text{ J}$$

Example 28-9 Confirm for the special case of an ideal solenoid, of N turns, length l, and radius a, that the ***magnetic energy per unit volume*** u_M may be expressed in the form

$$u_M = \frac{B^2}{2\mu_0} \tag{28-15}$$

Solution Since the **B**-field inside an ideal solenoid is $\mu_0 Ni/l$, (28-14) may be expressed as

$$U_M = \frac{1}{2\mu_0}\left(\frac{\mu_0 Ni}{l}\right)^2 l\pi a^2 = \frac{B^2 l\pi a^2}{2\mu_0}$$

and since the volume of the solenoid is $l\pi a^2$, (28-15) follows directly. Although here derived only for the ideal solenoid, (28-15) is found to be valid very generally.

28-6 Oscillations of the *L-C* circuit

An interesting illustration of the notion of electromagnetic energy deals with the discharge of a capacitor through an inductor. Consider, in Figure 28-10, a capacitor that has an initial charge Q_0 and suppose that at $t = 0$ it is connected across an inductor L. Although there is inevitably some resistance in any circuit, for the moment let us neglect it. If $i = i(t)$ represents the current in the circuit at any time t and $q(t)$ is the charge on the capacitor at this instant, then, the circuit equation is

$$\frac{1}{C}q = -L\frac{di}{dt} \tag{28-16}$$

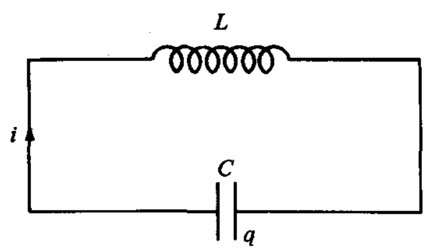


Figure 28-10

Furthermore, assuming that the right-hand plate of the capacitor has its charge q, the current is dq/dt, and therefore the circuit equation may be expressed equivalently as

$$\frac{d^2q}{dt^2} + \frac{1}{LC}q = 0 \tag{28-17}$$

The solution of this relation, subject to the initial conditions $q(0) = Q_0$ and $i(0) = 0$, gives a formula for the charge on the capacitor at any time t.

To solve (28-17), let us recall the equation of motion for the simple harmonic oscillator in (6-20); that is,

$$\frac{d^2x}{dt^2} = -\omega^2 x \tag{6-20}$$

and its solution in terms of the constants of integration A and α:

$$x(t) = A\cos(\omega t + \alpha) \tag{6-21}$$

Comparison of (28-17) with (6-20) and (6-21) shows then that the solution of (28-17) is

$$q(t) = A\cos\left(\frac{t}{\sqrt{LC}} + \alpha\right)$$

where, just as in (6-21), A and α are integration constants. Imposing the initial conditions $q(0) = Q_0$ and $i(0) = 0$, we find that $A = Q_0$ and $\alpha = 0$. The solution of (28-17) is therefore

$$q(t) = Q_0\cos\omega t \tag{28-18}$$

with ω, the *angular frequency of the oscillations*, defined by

$$\omega = \frac{1}{\sqrt{LC}} \tag{28-19}$$

Differentiating (28-18), we obtain a formula for the current in the circuit:

$$i = \frac{dq}{dt} = -\omega Q_0 \sin \omega t \tag{28-20}$$

The minus sign here is consistent with the fact that for $\omega t < \pi/2$, the capacitor discharges, and thus the direction of the current in the circuit is opposite to that assumed in the figure. (Recall that the right-hand plate of the capacitor was originally positive.)

The solutions in (28-18) through (28-20) show that the charge on the capacitor varies sinusoidally in time with period $P = 2\pi/\omega = 2\pi(LC)^{1/2}$. Immediately after being connected, the capacitor starts to discharge and continues to do so until, at time $\pi/2\omega$, $q = 0$. Thereafter its plates acquire charges of the opposite polarity, after which it discharges again, and then finally it regains its original charge Q_0. The cycle is then repeated. During this charging and discharging process, the current through the circuit also varies in time and increases, decreases, and reverses itself at the appropriate times. The current achieves its maximum value (ωQ_0) or minimum value ($-\omega Q_0$) when the charge on the capacitor is zero, and vanishes when the charge on the capacitor is $\pm Q_0$.

It is of considerable interest to examine the results in (28-18) and (28-20) from the viewpoint of electric and magnetic energy. The electrical energy U_E stored in the capacitor at time t is $q^2/2C$, and thus

$$U_E = \frac{1}{2C} q^2 = \frac{Q_0^2}{2C} \cos^2 \omega t \tag{28-21}$$

Similarly, the magnetic energy U_M stored in this stystem at time t is $\frac{1}{2}Li^2$, and hence

$$U_M = \frac{1}{2} Li^2 = \frac{1}{2} L\, \omega^2 Q_0^2 \sin^2 \omega t \tag{28-22}$$

$$= \frac{Q_0^2}{2C} \sin^2 \omega t$$

where the final equality follows by use of the definition for ω in (28-19). Note that, in accordance with conservation of energy ideas, the total energy $U = U_E + U_M$ is a time-independent constant with the expected initial value $Q_0^2/2C$. (Recall that $\sin^2 \omega t + \cos^2 \omega t = 1$.) Figure 28-11 shows a plot of these formulas for U_E and U_M as functions of time. At $t = 0$, $U_M = 0$ and U_E has its maximum value $Q_0^2/2C$. As time goes on, U_E decreases as the capacitor discharges, while the magnetic energy U_M rises with the current. At the instant $t = \pi/2\omega$ the capacitor is completely discharged, and thus U_E vanishes, whereas the current achieves a maximum. The total energy U is

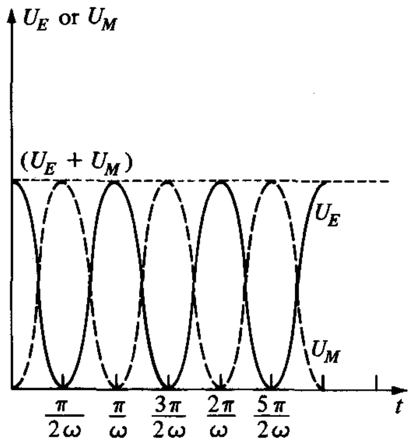


Figure 28-11

then all magnetic. After this the capacitor starts to be charged up with the opposite polarity until, at time $t = \pi/\omega$, the current falls to zero and the total energy is again stored entirely in the electric field of the capacitor. In this way, then, the total energy oscillates back and forth between the magnetic energy in the inductor and the electrical energy in the capacitor.

Example 28-10 A 0.1-μF capacitor has an initial charge of 2.0 μC and is connected across a 0.1-henry inductor. Calculate:

(a) The total energy in the system.
(b) The period of oscillation.
(c) The maximum current flow in the circuit.

Solution

(a) According to (28-21), the maximum energy stored in the capacitor U_E is

$$U_E = \frac{Q_0^2}{2C} = \frac{(2.0\times10^{-6}\,\mathrm{C})^2}{2\times(1.0\times10^{-7}\,\mathrm{F})}$$
$$= 2.0\times10^{-5}\,\mathrm{J}$$

(b) Since the period P of the oscillation is $2\pi/\omega$, it follows, from (28-19), that

$$P = \frac{2\pi}{\omega} = 2\pi\sqrt{LC} = 2\pi[(0.1\,\mathrm{H})\times(1.0\times10^{-7}\,\mathrm{F})]^{1/2}$$
$$= 6.3\times10^{-4}\,\mathrm{s}$$

(c) According to (28-20), the maximum current i_{max} is ωQ_0. Hence

$$i_{max} = \omega Q_0 = \frac{Q_0}{\sqrt{LC}} = \frac{2.0\times10^{-6}\,\mathrm{C}}{[0.1\,\mathrm{H}\times1.0\times10^{-7}\,\mathrm{F}]^{1/2}}$$
$$= 0.02\,\mathrm{A}$$

Example 28-11 Suppose that a resistor R is connected in series with the other elements of an L-C circuit. Assuming again that the capacitor has an original charge Q_0, find the charge on the capacitor at time t.

Solution In place of (28-17), the circuit equation now is

$$L\frac{d^2q}{dt^2}+R\frac{dq}{dt}+\frac{1}{C}q=0$$

since the potential drop across the resistor is Ri. Comparison with the equation of motion for a damped harmonic oscillator in (6-35) and its solution in (6-37) gives the solution

$$q(t)=Ae^{-Rt/2L}\cos(\omega' t+\alpha)$$

where, as in (6-37), A and α are integration constants. The parameter ω' is defined by

$$\omega'=\left[\frac{1}{LC}-\left(\frac{R}{2L}\right)^2\right]^{1/2}$$

and is real, if R is assumed to be small. It is left as an exercise to verify the validity of this solution and to calculate the values of A and α in terms of the initial conditions $q(0)=Q_0$ and $i(0)=0$.

28-7 Diamagnetism, paramagnetism, and ferromagnetism

In studying dielectrics we found it convenient to make use of the uniform electric field produced in a parallel-plate capacitor. In a similar way, in order to study the magnetic properties of matter it is convenient to make use of the uniform magnetic induction **B** that exists in the interior of an ideal solenoid. In practice, measurements in the laboratory are usually carried out by use of a toroid. However, since the **B**-field in a toroid is not uniform it is much simpler conceptually to carry out the analysis by use of a solenoid.

Suppose that a very long ideal solenoid has a length l and a radius a, and consists of N turns. A measurement of the self-inductance L_0 of this coil leads, in accordance with (28-4), to the value

$$L_0=\frac{\mu_0 N^2\pi a^2}{l} \tag{28-23}$$

where the subscript on L_0 is to emphasize the fact that this formula is valid provided that no magnetizable matter is nearby.

Consider again this solenoid, but suppose now that it is wound on a long, homogeneous cylinder of matter of radius a and of length l. A measurement of the self-inductance of the coil under this circumstance yields a new value, L, which, in general, will differ from the value L_0. However, further experimentation shows that the dependency of L on the parameters N, l, and a is precisely that given in (28-23). In other words, L differs from L_0 only by an overall factor. By analogy to the definition of the dielectric constant κ in Section 22-8, we define the *relative permeability* κ_m of the sample by

$$L=\kappa_m L_0 \tag{28-24}$$

so that κ_m is the ratio of the inductances of the coil with and without matter being present. As will be seen in the following section, physically the parameter κ_m represents the ratio B/B_0, where, for a given current i in the coil, B is the **B**-field in the sample, and B_0 is the **B**-field in the solenoid after the sample is removed. A related parameter, χ_m, is also frequently used to characterize magnetic materials. It is known as the *magnetic susceptibility* and is defined by

$$\chi_m = \kappa_m - 1 \tag{28-25}$$

Note that both κ_m and χ_m are dimensionless.

Before turning to a physical interpretation of magnetic susceptibility in the next section, let us summarize briefly the experimental situation. For the case of dielectrics, we found that the electric susceptibility χ_e is always positive. For magnetic materials the situation is not quite so simple. Experiments show that both positive and negative values for χ_m occur in nature.

It is convenient to classify magnetic substances into three main categories: *diamagnets*, *paramagnets*, and *ferromagnets*.[1]

A *diamagnetic material* is one whose magnetic susceptibility χ_m is very small and *negative*. A typical value is of the order of -10^{-5}. Table 28-1 lists (static) values of χ_m for a number of diamagnetic materials at room temperature. By contrast, a *paramagnetic material* is one for which the magnetic susceptibility, although still very small in magnitude, has a *positive* numerical value. In Table 28-1 values for χ_m for several paramagnetic substances are also listed.

Table 28-1 Magnetic susceptibilities at room temperature

Diamagnetic materials		*Paramagnetic materials*	
Substance	χ_m	Substance	χ_m
Bi	-1.7×10^{-5}	Al	2.5×10^{-5}
Cd	-2.3×10^{-6}	Ca	1.4×10^{-5}
C (diamond)	-2.2×10^{-5}	O_2 (gas)	1.8×10^{-6}
Cu	-1.1×10^{-6}	Pt	2.0×10^{-4}
Hg	-2.4×10^{-5}	Th	1.6×10^{-6}
Pb	-1.6×10^{-5}	W	3.5×10^{-6}

The third class of magnetic materials of particular interest consists of *ferromagnets*. Generally speaking, at room temperature most substances are either diamagnetic or paramagnetic. For these the magnetic susceptibility is generally independent of the strength of the current in the coils used to produce the magnetic behavior. By contrast, the elements Fe, Co, Ni, Gd, and Dy, and some of their alloys and oxides, are ferromagnetic and exhibit an altogether different behavior. On measuring the inductance of a coil

[1]A more complete listing would also include ferrimagnets and antiferromagnets.

wound around a ferromagnet, we find for χ_m large positive values, typically of the order of several hundred to several thousand. That is, the magnetic susceptibility of ferromagnets is 7 to 8 orders of magnitude larger than that for paramagnetic substances! Furthermore, not only do the values of χ_m, as so determined, vary with temperature, but they also depend on the strength of the current in the coils. In addition, as will be discussed in the next section, ferromagnets exhibit a phenomenon known as *hysteresis*, one of whose manifestations is a dependency of χ_m on the history of the specimen. The fact that χ_m for a ferromagnet varies with the current in the coils implies, for example, that the inductance L of such an inductor depends not only on the geometric parameters characterizing it but on the current through it as well.

To sum up, then, by measuring the self-inductance of a coil wrapped around a sample of material we may ascertain its magnetic susceptibility χ_m in accordance with (28-24) and (28-25). Materials for which χ_m has a small negative value (independent of the current) are known as diamagnets. Paramagnets are identical to diamagnets in many of their observable properties except for the fact that χ_m is small and positive. Finally, ferromagnets are materials for which χ_m is very large ($\sim 10^3$) and positive. For a ferromagnetic material χ_m varies, in general, with the strength of the current through the surrounding coils. Moreover, associated with each ferromagnetic material there is a critical temperature, known as the *Curie-point*, above which the material becomes paramagnetic. For example, the Curie point of iron is 1043 K. This means that if iron is heated to a temperature above 1043 K it loses all of its ferromagnetic properties and behaves as an ordinary paramagnetic substance. Similarly, the ferromagnetic element gadolinium (Gd) becomes paramagnetic at 289 K.

28-8 Magnetization currents

Consider a long, ideal solenoid of N turns and length l, and carrying a current i. According to (28-2), the magnetic flux (assuming no magnetizable bodies nearby) Φ_m° through a single turn of a coil is

$$\Phi_m^\circ = \frac{L_0 i}{N} \tag{28-26}$$

with L_0 the inductance of the coil.

Consider again this same coil, but suppose this time it is wrapped around a cylinder of matter of magnetic susceptibility χ_m (see Figure 28-12). The flux Φ_m through a single turn of the coil is now

$$\begin{aligned}\Phi_m &= \frac{Li}{N} = \frac{(1+\chi_m)L_0 i}{N} \\ &= (1+\chi_m)\Phi_m^\circ \end{aligned} \tag{28-27}$$

where the second equality follows by use of (28-24) and (28-25), and the last from (28-26). Thus, for paramagnetic and ferromagnetic materials (for which $\chi_m > 0$), the magnetic flux Φ_m through the coil is increased whereas for diamagnetic materials (for which $\chi_m < 0$) it is decreased. It is convenient to think of this increase—or decrease, for the case of a diamagnetic substance—of magnetic flux through the sample in terms of a *magnetization current* i_m flowing on its surface. Figure 28-12 illustrates the case for $\chi_m > 0$, for which the direction of flow of i_m is along that of the current i through the coils. For diamagnetic materials for which $\Phi_m < \Phi_m^\circ$, the direction of this magnetization current is in the opposite direction. Note that in both cases the total **B**-field inside the sample is the vector sum of the fields produced by each of the currents: i and i_m. This way of characterizing **B**-fields inside matter in terms of i_m is particularly useful for ferromagnetic substances which may exhibit magnetic behavior even when the external current i vanishes.

To discuss the physical mechanism that underlies these magnetization currents, let us consider first a homogeneous paramagnetic substance. Studies in quantum mechanics have shown that for these, the constituent molecules may be thought of as small magnetic dipoles. That is, the molecules of a paramagnetic substance behave as if they consisted of small current loops, which under normal circumstances are oriented randomly, so there is no net magnetic effect due to them. Consider now, in Figure 28-13, a cross section of such a paramagnetic material inside a long solenoid around which flows a certain current i. The microscopic current loops in the sample find themselves in a certain magnetic induction field $\mathbf{B}_0$ and, in accordance with (26-24), they experience a torque, which tends to line them up with their normals parallel to $\mathbf{B}_0$. Reference to the figure shows that when they are lined up, these currents due to neighboring loops cancel in the interior of the sample. Hence there is no magnetization current in this region. However, this cancellation cannot take place at the surface. Effectively, therefore, a certain magnetization current i_m flows on the surface along the direction of the causal current i. The total **B**-field inside the sample is due to this total current $(i + i_m)$ on its surface.

With regard to ferromagnetic materials, the situation is very similar, although more complex in detail. If the sample is initially unmagnetized, then its microscopic current loops are oriented at random as for the case of a

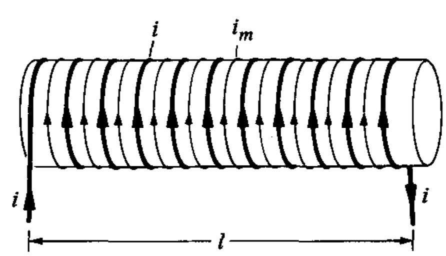


Figure 28-12

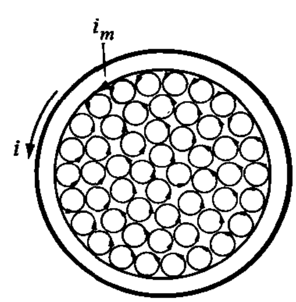


Figure 28-13

paramagnet. If the sample is then placed into an external **B**-field, these magnetic dipoles, or small current loops, again tend to line up with the external **B**-field, just as for the case of paramagnetic substance, but this time in a much more organized way. Indeed, for ferromagnetic materials, for which $\chi_m \gg 1$, the surface current i_m is very much greater than that in the surrounding coils, and thus the torque experienced by a given dipole is due mainly to the **B**-field produced by its neighbors. Moreover, the collective effect of these torques between the dipoles is so strong that for temperatures below the Curie point, they will remain lined up even after the external current is removed. From a macroscopic point of view, they thereby constitute a permanent magnet. This behavior is to be contrasted with that in a paramagnetic material, whose magnetization current vanishes as the causal current does.

To see the kinds of effects observed in laboratory experiments with ferromagnets consider, in Figure 28-14, a plot of the **B**-field inside a ferromagnetic sample as a function of the current i in the surrounding coils. Suppose that originally the sample is unmagnetized and is thus at the point A in the figure. As the current is turned on and gradually increased, the **B**-field inside the sample is found to increase steadily and to saturate at a maximum value at C. As the current is subsequently decreased, instead of retracing the original curve AC backward the **B**-field is found to follow along a different path CD, associated with consistently higher values for $|\mathbf{B}|$. In particular, even when the current through the coils vanishes, the **B**-field in the sample at point D does not! In other words, the sample has now become a *permanent magnet*. As noted above, once a ferromagnetic specimen has been magnetized, an external **B**-field is not required to keep the dipoles aligned; the fields produced by the individual dipoles are sufficient for this purpose. As the sense of the current around the sample is reversed, the **B**-field variations now follow curve DEF, and a second reversal of the current then leads to an increasing **B**-field along path FGH, and finally to saturation at C. Again, at point G there is a residual field, but now with a sense opposite to that at D. It

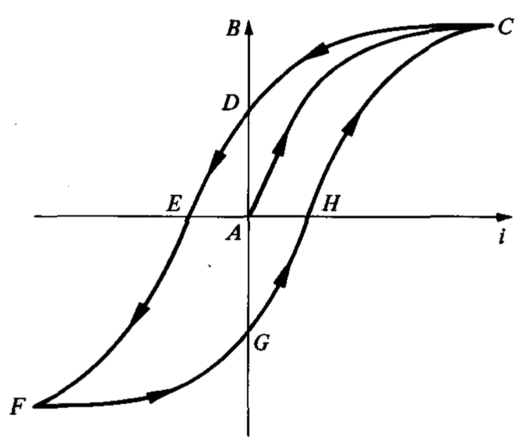


Figure 28-14

is because of these residual fields at D and G that one often says that the magnetic susceptibility of ferromagnets depend on their past history. The curve AC in the figure is usually referred to as the magnetization curve, and the closed loop $CDEFGHC$ is called a *hysteresis loop.*

Finally, let us consider diamagnetic substances, for which $\chi_m < 0$ and for which the magnetization current i_m flows in a direction opposite to the current in the coils. Consider, in Figure 28-15a, an electron in a circular orbit perpendicular to a uniform magnetic induction $\mathbf{B}_0$. Because of this **B**-field, the electron experiences the force $q\mathbf{v}\times\mathbf{B}$, and therefore, because of its negative charge it will orbit about a field line in the sense shown. However, because of its negative charge, the electron produces a circular current i with sense opposite to that of its velocity. Hence the magnetic induction $\mathbf{B}_e$ due to the current associated with this orbiting electron will be directed opposite to the causal field $\mathbf{B}_0$.

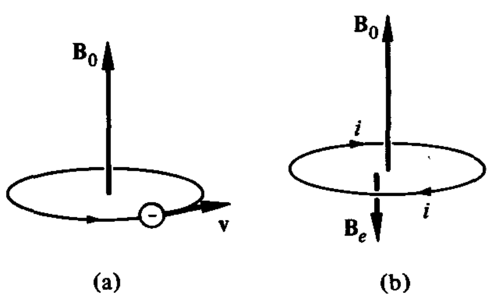


Figure 28-15

Consider now a sample of a diamagnetic material, such as bismuth, in a solenoid that carries a current i. The free electrons in their motions through the lattice will behave as small current loops just as for the case of paramagnets. But this time the magnetization current on the surface of the sample will be *opposite* to that of the external current i. Hence, the effective magnet flux through the sample is decreased and in accordance with (28-27) the magnetic susceptibility for diamagnetic materials is negative.

To summarize, then, in a paramagnetic material the constituent molecules can be thought of as small current loops, which in response to the torque produced by an external **B**-field tend to line up along that field. A ferromagnet is one for which this lining up is very much more organized than for a paramagnet. Finally, in a diamagnet the electrons can be thought of as orbiting in circles about the lines of magnetic induction, thereby producing a **B**-field directed opposite to the external one. It should be noted that more detailed studies show that most substances have diamagnetic as well as paramagnetic properties and that the observed magnetic behavior of any substance depends on which one of these dominates.

†28-9 The magnetization vector

In discussing dielectrics we found it convenient to define a vector **P** to represent the number of electric dipoles per unit volume in the sample. In a similar way, for magnetic materials it is convenient to define a *magnetization vector* **M** as the number of magnetic dipoles per unit volume in the material.

Consider, in Figure 28-16, a region of space in which there are magnetized bodies. Let l be an arbitrary, directed curve in this region, and let i_m represent the net magnetization current, which flows along the positive sense (defined by the right-hand rule) through an open surface bounded by l. The curve l itself may be entirely inside the material or partially inside and partially outside. In terms of i_m, define the *magnetization vector* **M** associated with this material by the requirement

$$\oint_l \mathbf{M} \cdot d\mathbf{l} = i_m \tag{28-28}$$

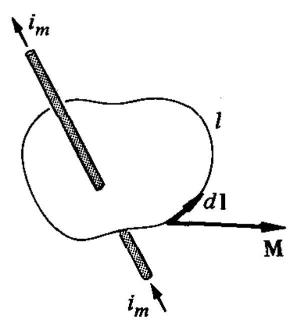


Figure 28-16

where i_m is the magnetization current flowing through any open surface bounded by the curve l, and the line integral is around l. It is implicit here that this relation is valid for *all* closed curves l. In the space outside of material bodies $i_m = 0$, and, consistent with (28-28), we shall assume that $\mathbf{M} \equiv 0$ in such regions. For the same reason that Ampère's law, (25-12), does not, in general, associate a unique **B**-field with a given current distribution, (28-28) does not determine a unique **M** for a given i_m either. Nevertheless, this definition suffices for our needs.

To determine the relation between **M** and the magnetic susceptibility χ_m of a material consider again, as in Figure 28-12, an ideal solenoid wrapped around a magnetic substance. If a current i is sent through the coil, a certain magnetization current i_m will flow on the surface of the sample. Let us define the *magnetic field* **H** inside the sample so that it satisfies the relation

$$\oint_l \mathbf{H} \cdot d\mathbf{l} = i \tag{28-29}$$

where, as in Ampère's law, the line integral is around a closed curve and i is the net current flowing through an open surface S bounded by that curve, with the positive sense given by the right-hand rule. To understand the

significance of **H**, let us compare (28-29) with Ampère's law,

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0(i + i_m) \tag{28-30}$$

where the form of the right-hand side follows since the *total* current flowing through a surface bounded by l is $(i + i_m)$. Comparing this with (28-29), we see that the magnetic field **H** for the situation in Figure 28-12 is the same as the product of μ_0^{-1} *and the* **B**-*field in the absence of magnetization current.* In other words, if no matter is present, so that $i_m = 0$, then up to a factor μ_0 the two fields **H** and **B** are identical.

There is a simple algebraic relation between **B**, **H**, and **M**, which may be obtained as follows. Substituting for the currents i and i_m in (28-30) by use of (28-28) and (28-29) and assuming the same curve l in all three integrals, we obtain

$$\begin{aligned}\oint \mathbf{B} \cdot d\mathbf{l} &= \mu_0(i + i_m)\\ &= \mu_0\left\{\oint \mathbf{H} \cdot d\mathbf{l} + \oint \mathbf{M} \cdot d\mathbf{l}\right\}\end{aligned}$$

Hence, since the curve l is arbitrary, it follows that[2]

$$\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) \tag{28-31}$$

Although derived only for the system in Figure 28-12, this relation is applicable very generally and is basic to all discussions of the magnetic behavior of matter.

We shall now derive the relation between **M** and χ_m. For the situation in Figure 28-12 the flux Φ_m through a single turn of the coil is

$$\Phi_m = \int_S \mathbf{B} \cdot d\mathbf{S}$$

where the integral goes over a surface bounded by a turn. Correspondingly, the magnetic flux Φ_m° through a single turn of the coil in the absence of the sample is

$$\Phi_m^\circ = \int_S \mu_0\mathbf{H} \cdot d\mathbf{S}$$

since the magnetic induction **B** is the same as $\mu_0\mathbf{H}$ in the absence of matter. Substituting these formulas for Φ_m and Φ_m° into (28-27), we find, assuming a "linear material" for which χ_m is independent of current and therefore of **H**, that

$$\int_S \mathbf{B} \cdot d\mathbf{S} = \mu_0(1 + \chi_m) \int_S \mathbf{H} \cdot d\mathbf{S}$$

[2]Strictly speaking, since $\oint d\mathbf{l} = 0$ for any closed curve l (Example 26-7), this argument only shows that the vector $[\mathbf{B} - \mu_0(\mathbf{H} + \mathbf{M})]$ is a constant. However, since $\mathbf{M} = 0$ outside of matter, and since for the ideal solenoid $\mathbf{H} = \mathbf{B} = 0$ outside, it follows that this constant is zero.

Since the surface S is arbitrary, this is equivalent to

$$\mathbf{B} = \mu_0(1 + \chi_m)\mathbf{H} \tag{28-32}$$

and combining this with (28-31), we obtain the sought-for formula

$$\mathbf{M} = \chi_m \mathbf{H} \tag{28-33}$$

for the magnetization $\mathbf{M}$ inside the sample.

The value for χ_m, as we saw above, can be determined by a measurement of inductance. Hence, since $\mathbf{H}$ is known independently it follows that $\mathbf{M}$ itself can be measured. The definition in (28-28) then enables us to calculate the magnetization currents i_m associated with the given sample.

Example 28-12 Suppose that a current of 2.0 amperes flows in the coils in Figure 28-12 and that the sample is iron, for which $\chi_m = 200$. Assuming that there are 10^4 turns per meter and that (28-32) and (28-33) are applicable, calculate:

(a) The magnetic field strength $\mathbf{H}$ in the sample.
(b) The magnetization vector $\mathbf{M}$.
(c) The magnetic induction $\mathbf{B}$.

Solution

(a) In the absence of the sample $\mathbf{B} = \mu_0\mathbf{H}$. Since $\mathbf{B}$ is uniform and parallel to the axis of the coil and has a strength $\mu_0 ni$, it follows that the magnitude of the magnetic field is

$$H = ni = (10^4/\text{m}) \times 2.0\ \text{A} = 2.0 \times 10^4\ \text{A/m}$$

(b) Using the given value $\chi_m = 200$ and the above value for H, we find by use of (28-33) that

$$\begin{aligned} M &= \chi_m H = (200) \times (2.0 \times 10^4\ \text{A/m}) \\ &= 4.0 \times 10^6\ \text{A/m} \end{aligned}$$

(c) The substitution of these values into (28-32) leads to

$$B = \mu_0(1 + \chi_m)H = (4\pi \times 10^{-7}\ \text{T-m/A}) \times (201) \times (2.0 \times 10^4\ \text{A/m}) = 5.0\ \text{T}$$

The directions of $\mathbf{B}$, $\mathbf{H}$, and $\mathbf{M}$ are all parallel to the axis of the coil, and with their sense given by the right-hand rule.

It is interesting to note that the effect of the ferromagnetic core here is to increase the $\mathbf{B}$-field inside the solenoid by the factor $(1 + \chi_m) \approx 200$. The iron core thus effectively decreases the current required to produce a given $\mathbf{B}$-field by a factor $(1 + \chi_m)$. Equivalently, the magnetic induction $\mu_0\mathbf{H}$ produced by the external current i is negligible compared to the $\mathbf{B}$-field inside the sample. Thus, we can think of a material characterized by a very large χ_m as one which tends to pull $\mathbf{B}$-lines into its interior. This feature plays an important role in variety of applications; for example, in the design of galvanometers, as we saw in Section 26-7.

28-10 Summary of important formulas

The self-inductance L of a loop is defined by

$$\Phi_m = Li \tag{28-1}$$

where Φ_m is the total flux through the loop when it carries a current i. The magnetic energy U_M stored in the loop when it carries a current i is

$$U_M = \frac{Li^2}{2} \tag{28-13}$$

If L_0 is the self-inductance of a coil, then the inductance L of the same coil when wound around a homogeneous sample is given by

$$L = \kappa_m L_0 \tag{28-24}$$

where κ_m is the relative permeability of the sample.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** mutual inductance; **(b)** relative permeability; **(c)** magnetic susceptibility; and **(d)** magnetization vector.
2. What important feature (or features) distinguishes diamagnetic, paramagnetic, and ferromagnetic materials from each other?
3. Making use of (28-3), explain why the rules for combining inductors in series and parallel are the same as those for combining resistors.
4. Two identical loops of wire, each of one turn, are placed near each other so that they almost coincide. If L is the self-inductance of either loop, why must the mutual inductance of the pair also have the value L?
5. Associated with any real circuit there is, in general, some resistance. Explain in physical terms why it is that associated with any circuit there is also inevitably some inductance. For what type of circuits would this effect be a serious problem?
6. Is it possible for the self-inductance of a coil to have a negative value? What about the mutual inductance of two coils? Explain.
7. Explain in physical terms why and under what circumstances an inductor in a circuit behaves as an infinite resistor. Does the magnetic energy $Li^2/2$ stored in the inductor have anything to do with this feature?
8. Explain in terms of magnetic energy why and under what circumstances an inductor in a circuit behaves as a short; that is, as a resistor with zero resistance.
9. In more advanced studies, it is shown that a formula for the mutual inductance M between two coils is

$$M = \mu_0 \oint_{l_1} \oint_{l_2} \frac{d\mathbf{l}_1 \cdot d\mathbf{l}_2}{r}$$

where the symbols are as defined in Figure 28-17. Show by use of this formula that the reciprocity law in (28-7) is valid.

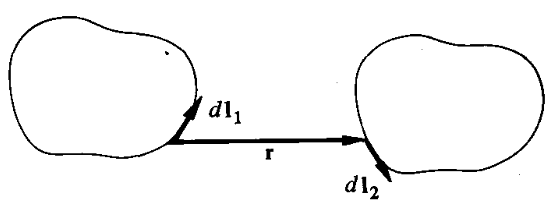


Figure 28-17

10. Consider the formula in (28-4) for the self-inductance of an ideal solenoid. Consider the limiting case $N \to \infty$, $l \to \infty$, with the ratio $N/l = n$ fixed. Give a physical interpretation for the fact that L diverges in this limit, and explain why this is not a practical problem.
11. Explain in terms of magnetic energy, why it is not possible to alter suddenly the current in a circuit that contains an inductance. Can you

think of a practical use for this feature?

12. Suppose that an ammeter were inserted in the R-L circuit in Figure 28-6. Describe the motion of its pointer as a function of time if **(a)** $\tau_L = 1$ ms; **(b)** $\tau_L = 5$ seconds. (Assume that a full-scale reading on the ammeter corresponds to the maximum current in the circuit.)
13. If a voltmeter is connected across the inductor in Figure 28-6, describe the motion of the pointer if **(a)** $\tau_L = 1$ ms; **(b)** $\tau_L = 5$ seconds. (Assume that $\mathscr{E}$ is the full-scale deflection on the voltmeter.)
14. An unmagnetized iron nail is brought near a permanent magnet. Why will the nail be attracted to the magnet? If a second nail is brought near the first, what happens?
15. Explain in microscopic terms the distinction between paramagnetism and diamagnetism. What would happen to a diamagnetic needle suspended in a magnetic field? Contrast this behavior with that of a paramagnetic needle.
16. Consider a magnetized iron rod. Explain in microscopic terms why, if this rod is cut in half, a new north and south pole appear at the cut surfaces. Could an isolated magnetic pole be obtained in this way? Explain.
17. Because isolated magnetic poles do not exist, the flux of **B** out of any closed surface vanishes. Why is this not true, in general, for the flux of **H** out of a closed surface? Under what circumstances is the relation

$$\oint_S \mathbf{H} \cdot d\mathbf{S} = 0$$

true?

18. What happens to a ferromagnet when its temperature is raised above the Curie point?
19. Suppose a long, thin rod is inserted into an inductor in an R-L circuit. If the time constant of the circuit is found to increase slightly, is the rod diamagnetic, ferromagnetic, or paramagnetic? Explain.
20. A coil of wire carrying a current i is wrapped around a long, iron rod. Explain why most of the lines of magnetic induction will be in the interior of the rod even if the coil is of relatively short length.

PROBLEMS

1. A long, ideal solenoid is 2 cm in diameter and has 40 turns/cm.
   (a) What is the self-inductance per meter of this coil?
   (b) If a current of 1.5 amperes flows through the coil, what is the magnetic flux in a single turn?
   (c) What is the total flux through the turns in 0.5 meter of the coil?
2. A long, ideal solenoid has 2×10^3 turns and an inductance of 2.0 henry.
   (a) What is the total magnetic flux through the solenoid if a current of 0.3 ampere flows?
   (b) If the solenoid has a cross sectional area of 2 cm^2, what is the strength of the magnetic induction inside the coil when this current flows?
3. A toroid has a square cross section, an inner radius of 10 cm, and an outer radius of 12 cm. If it has 300 turns:
   (a) Calculate the self-inductance.
   (b) Calculate the magnetic flux through a single turn if a steady current of 3.0 amperes flows.

***4.** A toroid of inner radius a and outer radius b has a circular cross section of radius $(b-a)/2$ and N turns. Calculate the self-inductance L of this coil. (*Hint*: Use the method in Example 28-2.)

5. Consider the formula for the self-inductance of a toroid in (28-5). Show that in the limit as the radius of the toroid becomes very large (while its cross-sectional area is fixed) the formula for its inductance per unit length reduces to that for a very long, straight solenoid.

6. A coil of 10 turns is wrapped around a very long, ideal solenoid of self-inductance 5.0 mH and of 10^3 turns.
- **(a)** If a current i flows in the solenoid, what is the magnetic flux through a single turn of the *coil*?
- **(b)** What is the mutual inductance between the coil and the solenoid?

7. A coil of 100 turns is wrapped around a long, ideal solenoid of self-inductance 2.0 henry and having 20,000 turns.
- **(a)** What is the mutual inductance between the coil and the solenoid?
- **(b)** If a current of 10 amperes flows in the coil, what is the magnetic flux through a single turn of the solenoid? What is the total flux through the entire solenoid?

8. Suppose the toroid in Figure 28-18 has 100 turns, an inner radius of 20 cm, an outer radius of 25 cm, and a square cross section.
- **(a)** What is the self-inductance of the toroid?
- **(b)** What is the mutual inductance between the toroid and the loop?
- **(c)** If a current of 2.0×10^{-2} ampere flows in the toroid what is the magnetic flux through the loop?

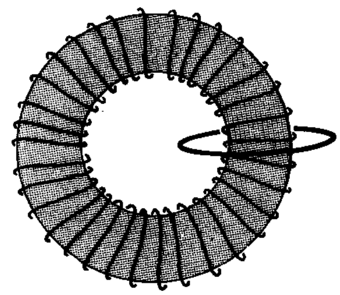


Figure 28-18

***9.** Consider the situation in Figure 28-5. If a current i flows in the rectangular loop, what is the total magnetic flux through the area bounded by the two parallel wires? Assume that there is no current flowing in the long wires. (*Hint*: Use the formula for M in Example 28-4.)

10. A 10-volt battery is suddenly connected across a 10-mH inductor and a 100-Ω resistor in series.
- **(a)** What is the initial current in the circuit?
- **(b)** What is the initial potential across the inductor?
- **(c)** What is the initial value for the time rate of change of the current?

11. Consider the same circuit as in Problem 10.
- **(a)** What is the time constant associated with this circuit?
- **(b)** What current flows in the circuit at the instant $t = \tau_L$?
- **(c)** What is the potential drop across the resistor at $t = 10\tau_L$?

12. What value for the inductance is required in an R-L circuit so that for a resistance of 75 Ω the time constant will be 2.0 ms? If the emf of the battery is 100 volts, how much energy is stored in the inductor under steady-state conditions?

13. A certain R-L circuit has a time constant τ_L. Show that the time $t_{1/2}$ that must elapse for the current to achieve half of its final steady-state value is $t_{1/2} = \tau_L \ln 2$.

14. A man holds in his right hand a long, ideal solenoid of self-inductance 0.2 henry, around which flows the current $i = \alpha + \beta t$, with $\alpha = 10$ amperes and $\beta = -10$ A/s. Suppose that for $t \leq 1$ second, the direction of the current through the coils is along his fingers.
   (a) What is the magnitude of the emf induced in the solenoid at $t = 0.5$ second?
   (b) Does the man's thumb point in the direction of increasing or decreasing potential?
15. Two inductors L_1 and L_2 are very far away from each other.
   (a) Show that the equivalent inductance L if they are connected in series is $L = L_1 + L_2$.
   (b) Show that the equivalent inductance if they are in parallel is

$$L = \frac{L_1 L_2}{L_1 + L_2}$$

16. Consider the circuits in Figure 28-19 and suppose that M is the mutual inductance between the inductors. Show that the circuit equations are:

$$R_1 i_1 + L_1 \frac{di_1}{dt} + M \frac{di_2}{dt} = \mathscr{E}$$

$$R_2 i_2 + L_2 \frac{di_2}{dt} + M \frac{di_1}{dt} = 0$$

What are the values for i_1 and i_2 after a long time when steady-state conditions have been reached?

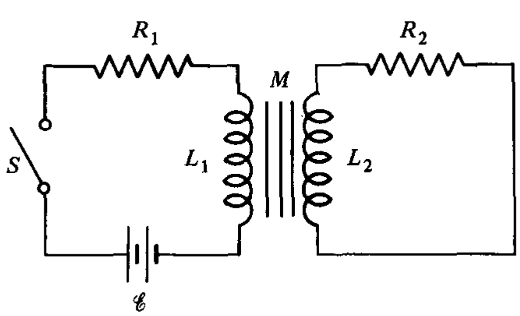


Figure 28-19

*17. Using the method of Section 28-5 and the results of Problem 16, verify that the magnetic energy U_M stored in the circuits in Figure 28-19 is

$$U_M = \frac{1}{2} L_1 i_1^2 + \frac{1}{2} L_2 i_2^2 + M i_1 i_2$$

18. Consider two coaxial ideal solenoids of radii a and b ($> a$), respectively. Suppose that the outer one has N_2 turns and the inner one N_1 turns, and that they have the same length l.
   (a) Show that the **B**-field at a distance r from the axis has the magnitude

$$B = \frac{\mu_0}{l} \begin{Bmatrix} (N_1 i_1 + N_2 i_2) & r \leq a \\ N_2 i_2 & a \leq r \leq b \\ 0 & r \geq b \end{Bmatrix}$$

   where i_1 and i_2 are the currents in the inner and outer coils, respectively.
   (b) Calculate the magnetic energy U_M by use of (28-15).
   (c) Using your result to (b) and Problem 17, show that the mutual inductance M of these solenoids agrees with that in Example 28-3.
19. If a voltmeter were connected across the inductor in Figure 28-6, show that at time t after the switch is closed it would read $\mathscr{E} e^{-t/\tau_L}$.
*20. Two inductors L_1 and L_2 are connected in series in a way so that their mutual inductance is M. Show that the inductance L of the equivalent inductor is

$$L = L_1 + L_2 + 2M$$

*21. If the two inductors of Problem 20 are connected in parallel, show that the equivalent inductance is

$$L = \frac{L_1 L_2 - M^2}{L_1 + L_2 - 2M}$$

22. Two inductors, each having a self-inductance of 0.4 henry, are connected in parallel across a 20-volt battery.

(a) If the final steady current through the battery is 2.0 amperes, what is the resistance in the circuit?

(b) If the R-L time constant is found to be 0.05 second, what is the equivalent inductance?

(c) Making use of the result of Problem 21, calculate the mutual inductance of the two coils.

23. Consider the circuit in Figure 28-20 and suppose that the switch S is closed at $t = 0$.

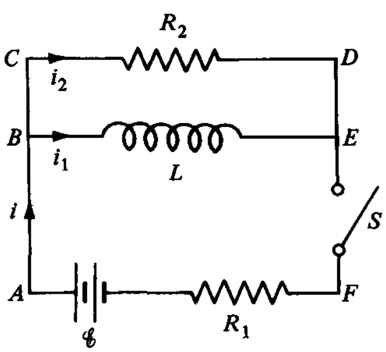


Figure 28-20

(a) Describe the path followed by the current immediately after the switch is closed.

(b) What is the path followed by the current after a long time?

24. Assuming in Figure 28-20 the parameter values $\mathscr{E} = 20$ volts, $R_1 = 10\,\Omega$, $R_2 = 50\,\Omega$, and $L = 50$ mH, calculate:

(a) The initial values for i, i_1, and i_2.

(b) The values for i, i_1, and i_2 after a long time.

(c) The maximum value for the magnetic energy stored in the inductor.

25. Consider the circuit in Figure 28-21 and suppose that the parameters have the values $\mathscr{E} = 50$ volts, $R_1 =$

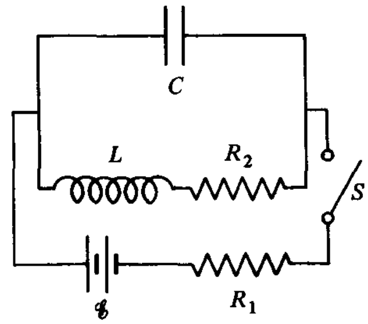


Figure 28-21

$50\,\Omega$, $R_2 = 10\,\Omega$, $L = 2.0$ henry, and $C = 5\,\mu$F. If the switch is closed at $t = 0$:

(a) What is the initial current through the battery?

(b) What is the current through the battery after a long time?

(c) What is the final charge on the capacitor?

26. For the circuit in Figure 28-21 calculate:

(a) The final value for the magnetic energy stored in the inductor.

(b) The final electric energy stored in the capacitor.

(c) The rate at which energy is dissipated in R_1 under steady-state conditions.

Use the parameter values in Problem 25.

27. For the circuit in Figure 28-22 assume the parameter values $\mathscr{E} =$ 40 volts, $R_1 = 10\,\Omega$, $R_2 = 60\,\Omega$, $R_3 = 30\,\Omega$, $R_4 = 20\,\Omega$, $C = 2\mu$F, $L_1 = 1.0$

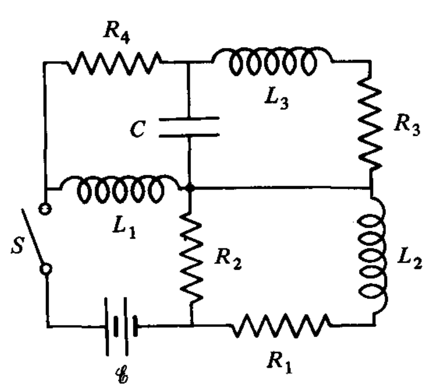


Figure 28-22

henry, $L_2 = 2.0$ henry, $L_3 = 3.0$ henry, and calculate:
(a) The initial value of the current through the battery.
(b) The final current through R_2.
(c) The final current through R_1.
(d) The final current through the battery.

28. For the physical situation described in Problem 27 calculate the final energy stored in each inductor and in the capacitor.

29. A capacitor of capacitance 2.0 μF has an initial charge of 5.0 μC and is connected across a 2.0-mH inductor.
(a) What is the angular frequency of the oscillations of this circuit?
(b) What is the period?
(c) What is the maximum energy stored in the inductor?

*30. A 2.0-μF capacitor with an original charge of 2.0 μC is connected across a 20-Ω resistor and a 2.0-mH inductance, as in Figure 28-23. By reference to the results of Example 28-11, calculate:

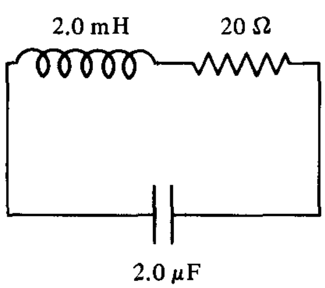


Figure 28-23

(a) The angular frequency of oscillations if the resistor were not present.
(b) The actual angular frequency ω'.
(c) The total energy dissipated in the resistor.

31. A time-dependent emf $\mathscr{E}\cos\omega_0 t$ ($\mathscr{E}$, ω_0 positive constants) is connected across an L-C combination, as in Figure 28-24.

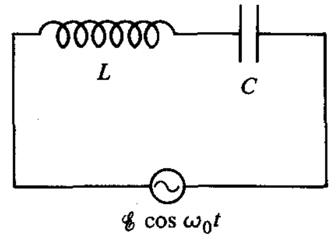


Figure 28-24

(a) Show that if $q(t)$ is the charge on the capacitor at time t, then the circuit equation may be written

$$L\frac{d^2q}{dt^2} + \frac{q}{C} = \mathscr{E}\cos\omega_0 t$$

(b) Verify that

$$q(t) = A\cos(\omega t + \phi) + \frac{(\cos\omega_0 t)\mathscr{E}/L}{(\omega^2 - \omega_0^2)}$$

where $\omega = 1/(LC)^{1/2}$ and A and ϕ are integration constants, is a solution of (a).
(c) Explain in physical terms what happens as the external frequency ω_0 becomes equal to the *resonant* frequency $(LC)^{-1/2}$ of this L-C combination.

*32. Consider again the circuit in Figure 28-24, but suppose that this time a resistor R is in series with the inductor.
(a) Write down the circuit equation in terms of the charge q on the capacitor and its time derivatives.
(b) Verify that a solution of this equation is

$$q(t) = \frac{\mathscr{E}\cos(\omega_0 t + \alpha)}{\left[\left(\omega_0^2 L - \frac{1}{C}\right)^2 + R^2\omega_0^2\right]^{1/2}}$$

where α satisfies

$$\sin\alpha = \frac{-R\omega_0}{\left[\left(\omega_0^2 L - \frac{1}{C}\right)^2 + R^2\omega_0^2\right]^{1/2}}$$

33. Making use of the results of Problem 32 for the particular choice of $L = 1.0$ mH and $C = 1.0\ \mu$F, make a plot of the maximum potential drop V_{max} across the resistor as a function of ω_0 for the following choices: **(a)** $R = 1\ \Omega$; **(b)** $R = 100\ \Omega$; and **(c)** $R = 10^3\ \Omega$. Explain by reference to your graphs why this circuit can be used as a frequency selector, for example, in a radio.

†34. Show by use of (28-28) that if **M** is the magnetization vector associated with a given magnetization current i_m, then $(\mathbf{M} + \mathbf{M}_0)$, where $\mathbf{M}_0$ is an *arbitrary* but constant vector, also satisfies this relation. How do we resolve this ambiguity in practice?

†35. Show by use of (28-32) that for a sample for which $\chi_m = 500$, the ratio of the strength of the magnetic induction inside the sample to that outside is approximately 500. It is for this reason that we often say that if a ferromagnetic material is introduced into a region containing a magnetic induction it tends to "suck" the field lines into its interior.

†36. A steel rod having a length of 0.5 meter has a relative permeability of 1000 and is initially unmagnetized. If we uniformly wrap around it 100 turns of wire, what current is required so that the magnetic induction inside the iron is 1.5 tesla? Calculate also the strength of **M** and **H** inside the iron.

†37. Figure 28-25 shows the basic elements of a *transformer*. N_1 turns of one coil, called the *primary* coil, are wound around a ring of high permeability such as iron. A coil, called the *secondary*, having N_2 turns is also wound about the ring.

(a) If a current i flows in the primary, why is it that the magnetic flux through a *single turn* of either coil is the same to a high degree of approximation.

(b) Show that if a time-dependent emf $\mathscr{E}$ is applied to the primary, then the voltage V across the secondary is

$$V = \frac{\mathscr{E} N_2}{N_1}$$

(*Note*: If $N_1 < N_2$, then $V > \mathscr{E}$, and the transformer is said to be a *step-up* transformer. Conversely, if $N_1 > N_2$, then it is a *step-down* transformer.)

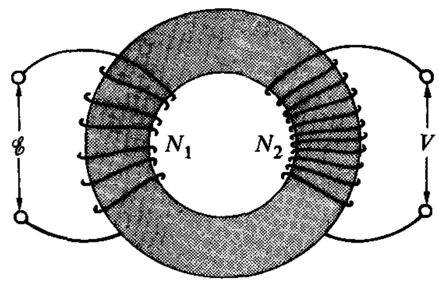


Figure 28-25

†38. **(a)** By constructing an appropriate Gaussian surface, part of which lies in the interior of magnetizable matter, show by use of Gauss' law for magnetism that the normal components of **B** are continuous across the interface.

(b) Show by use of Ampère's law that because of magnetization currents, the tangential components of **B** are *not* in general continuous across the surface of a magnetizable material.

†39. By arguments similar to those in Problem 38 show that the tangential components of **H** are continuous, but that the normal components of **H** may be discontinuous, across the interface between two magnetic materials.

29 Maxwell's equations and electromagnetic waves

One cannot escape the feeling that these mathematical formulas [Maxwell's equations] have an independent existence and an intelligence of their own, that they are wiser than we are, wiser even than their discoverers, that we get more out of them than we put into them.

HEINRICH R. HERTZ (1857–1894)

Who will then explain the explanation!

BYRON

29-1 Introduction

For the past ten chapters we have been studying the electric and magnetic fields produced by charges at rest and in motion. The important results of this study are the two laws of Gauss (one for electricity and one for magnetism) and the laws of Ampère and of Faraday. The laws governing specialized phenomena, such as those associated with electrostatics and magnetostatics, are all derivable from these four fundamental laws. Except for the addition to Ampère's law of a term known as the *displacement current* which we shall discuss in Section 29-3, these four relations comprise *Maxwell's equations.* When supplemented by certain empirical relations

such as Ohm's law, Maxwell's equations constitute a complete and quantitative description of *all* electromagnetic phenomena.

29-2 Recapitulation

As a preliminary to a discussion of Maxwell's equations and electromagnetic waves, in this section we review briefly the four basic relations of electromagnetism.

First, there is *Gauss' law* for electricity. This states that the total electric flux out of every closed surface S is the same as the product of $1/\epsilon_0$ and the *total* charge $q(S)$ *inside* that surface. In mathematical terms, then,

$$\oint_S \mathbf{E}\cdot d\mathbf{S} = \frac{1}{\epsilon_0} q(S) \qquad \text{(Gauss)} \qquad \mathbf{(29\text{-}1)}$$

where **E** is the electric field and, as in Figure 29-1a, the surface S is closed.

Because of the fact that magnetic monopoles—that is, isolated north or south magnetic poles—do not exist, the corresponding *Gauss' law for magnetism* is simpler and states that the total magnetic flux out of every closed surface vanishes. In mathematical terms, we may express this as

$$\oint_S \mathbf{B}\cdot d\mathbf{S} = 0 \qquad \text{(Gauss)} \qquad \mathbf{(29\text{-}2)}$$

where **B** is the magnetic induction and where again, as in Figure 29-1a, the surface S is presumed to be closed. Note that (29-2) is valid regardless of what currents or permanent magnets may be present inside or outside of the closed surface S.

A third important property of the electromagnetic field is given by *Faraday's law*. This states that the electromotive force developed about any closed path l is the same as the negative of the rate of change of magnetic

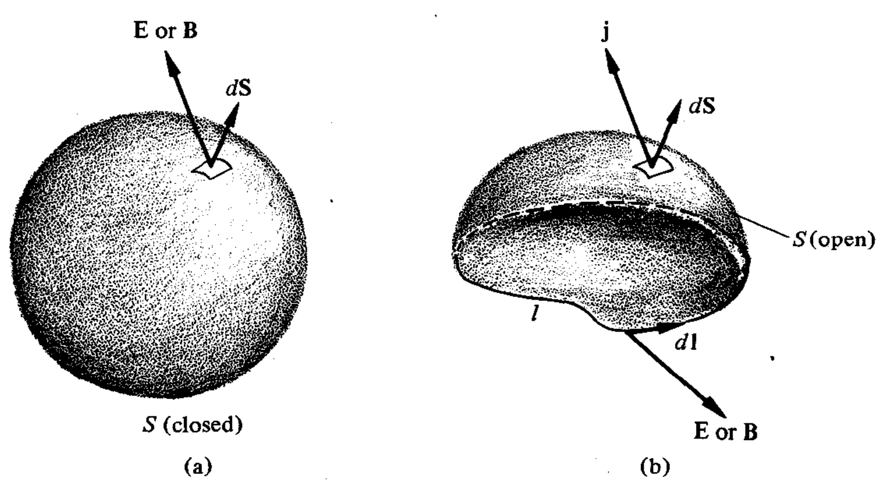


Figure 29-1

flux through every open surface S bounded by l. In the notation defined in Figure 29-1b, Faraday's law is

$$\oint_l \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt}\int_S \mathbf{B} \cdot d\mathbf{S} \qquad \text{(Faraday)} \qquad \textbf{(29-3)}$$

where the sense of the bounding curve and that of the outward normal to the surface S are related by the right-hand rule (Section 27-6).

Finally, there is *Ampère's law.* This may be expressed by the formula

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_S \mathbf{j} \cdot d\mathbf{S} \qquad \text{(Ampère)} \qquad \textbf{(29-4)}$$

where, as in Figure 29-1b, l is a directed, closed curve that bounds an arbitrary, but open, surface S. The current density $\mathbf{j}$ in this formula represents the vector sum of *all* currents, whether they be associated with the actual motion of charge or with magnetization currents inside or on the surface of matter. As for Faraday's law, the outward normal to S is assumed to be related to the sense of l by the right-hand rule.

It is worth emphasizing that all other relations, such as the equations of electrostatics or of magnetostatics or the law of Biot-Savart, are direct consequences of (29-1) through (29-4). For the phenomena of electrostatics, for example, $\mathbf{B}$ and $\mathbf{j}$ both vanish. Hence, (29-3) reduces for this case to

$$\oint_l \mathbf{E} \cdot d\mathbf{l} = 0$$

and the combination of this and Gauss' law, (29-1), constitute a complete specification of the electrostatic field. Similarly, the $\mathbf{B}$-field associated with a static current distribution may be calculated directly by use of (29-2) and (29-4). The law of Biot-Savart follows from these as a special case for currents in thin wires.

To summarize, then, the scope of the basic relations in (29-1) through (29-4) encompasses all phenomena of electromagnetism that we have studied up to now. However, these relations are *not* Maxwell's equations. Maxwell's equations result from (29-1) through (29-4) only after a certain inconsistency, associated mainly with Ampère's law, is removed.

29-3 Maxwell's equations

Maxwell was the first to point out that there is a difficulty with Ampère's law in (29-4) if it is assumed to be valid for *all* closed paths l. The details of his argument are somewhat complex and will be briefly touched upon in Section 29-10. In outline, Maxwell showed that Ampère's law in (29-4) leads to certain unphysical predictions if it is assumed to be valid for all closed curves. Specifically, he established that Ampère's law predicts that the net current flowing out of *every* closed surface must vanish. But it is easy to see

that this prediction is false. For if, as shown in Figure 29-2, one plate of a capacitor is surrounded by a closed surface S (the dashed surface in the figure), then during the time that the capacitor is being charged up, there is a definite flow of current into S. That is, since charge is accumulating on the plate of the capacitor, there must be a net current flow onto the plate and thus into S. Therefore, to be consistent with experiment, Maxwell reasoned, that Ampère's law in (29-4) cannot be correct as it stands.

Bearing this difficulty in mind, and making use of arguments of the type discussed in Section 29-10, Maxwell proposed adding the term

$$\epsilon_0\mu_0 \frac{d}{dt}\int_S \mathbf{E} \cdot d\mathbf{S}$$

to the right-hand side of Ampère's law. With this modification, Ampère's law becomes

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_S \mathbf{j} \cdot d\mathbf{S} + \mu_0\epsilon_0 \frac{d}{dt}\int_S \mathbf{E} \cdot d\mathbf{S} \tag{29-5}$$

where, as before, the sense of the surface S and its bounding curve l are related by the right-hand rule (Figure 29-1b). And it is this relation in (29-5) and the two laws of Gauss and that of Faraday in (29-1) through (29-3) that collectively are known as Maxwell's equations.

The above added term, which makes (29-5) differ from Ampère's law, is known as the *displacement current*, and it fell to Heinrich R. Hertz (1857–1894) to establish its undisputable physical reality by means of experiments involving electromagnetic waves. However, nearly two decades were to elapse from the time of Maxwell's original proposal until Hertz succeeded in demonstrating that—consistent with (29-5), Faraday's law, and the two laws of Gauss—electromagnetic waves do indeed exist. It is interesting to note in this connection that (29-1) through (29-4)—that is, Maxwell's equations without the displacement current—do *not* admit solutions corresponding to electromagnetic waves.

As noted earlier, our understanding of all electromagnetic phenomena

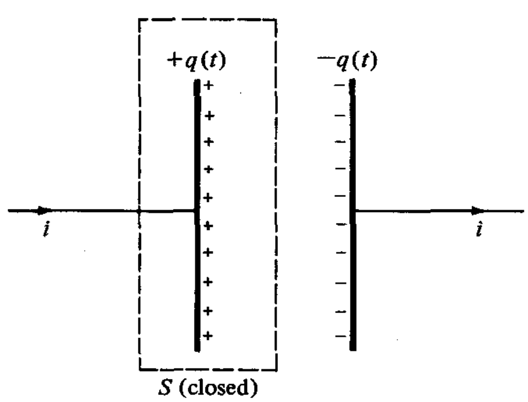


Figure 29-2

including those described in the preceding ten chapters is based on Maxwell's equations. Moreover, as will be shown in the following sections of this chapter, the existence and the properties of electromagnetic waves also follow from these relations. Maxwell's equations, in other words, describe not only the full range of ordinary electromagnetic phenomena, but also those of visible and ultraviolet light, radio waves, infrared radiation, X rays, and so forth. Hertz's observation that they "are wiser even than their discoverers" can hardly be classed as a metaphor.

Example 29-1 In the presence of material bodies characterized by a dipole moment per unit volume **P** and a magnetization **M**, Maxwell's equations consist of the two relations

$$\oint_S \mathbf{D} \cdot d\mathbf{S} = q_t(S)$$
$$\oint_l \mathbf{H} \cdot d\mathbf{l} = \int_S \mathbf{j}_t \cdot d\mathbf{S} + \frac{d}{dt}\int_S \mathbf{D} \cdot d\mathbf{S} \tag{29-6}$$

plus Gauss' law for magnetism, (29-2), and Faraday's law, (29-3). Here **D** is the displacement vector as defined in (22-26), **H** is the magnetic field as defined in (28-29) and q_t and $\mathbf{j}_t$ are all charges and currents except those induced in dielectric and magnetic materials. Express (29-6) directly in terms of **E** and **B**, assuming that S and l lie in the interior of a homogeneous, isotropic, and linear material.

Solution According to (22-31), for a linear material the displacement vector **D** is related to **E** via

$$\mathbf{D} = \kappa\epsilon_0\mathbf{E}$$

with κ the dielectric constant. Similarly, according to (28-25) and (28-32), **B** and **H** are related by

$$\mathbf{B} = \mu_0\kappa_m\mathbf{H}$$

with κ_m the relative permeability. Substitution into (29-6) leads to

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\kappa\epsilon_0} q_t(S)$$
$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0\kappa_m \int_S \mathbf{j}_t \cdot d\mathbf{S} + \epsilon_0\mu_0\kappa\kappa_m \frac{d}{dt}\int_S \mathbf{E} \cdot d\mathbf{S} \tag{29-7}$$

where κ and κ_m have been taken out from under the integral signs since they are constant. Note that these relations reduce to (29-1) and (29-5), respectively, if the parameters $\kappa\epsilon_0$ and $\kappa_m\mu_0$ are replaced by ϵ_0 and μ_0, respectively.

29-4 The solutions of Maxwell's equations in free space

Let us consider Maxwell's equations—(29-1), (29-2), (29-3), and (29-5)—in a region of space very far away from any charges and currents. For simplicity,

assume for the moment that no dielectrics or magnetizable bodies are nearby. Setting $q(S)=0$, $\mathbf{j}=0$, Maxwell's equations assume the form

$$
\begin{aligned}
&\oint_S \mathbf{E}\cdot d\mathbf{S}=0\\
&\oint_S \mathbf{B}\cdot d\mathbf{S}=0\\
&\oint_l \mathbf{E}\cdot d\mathbf{l}=-\frac{d}{dt}\int_S \mathbf{B}\cdot d\mathbf{S}\\
&\oint_l \mathbf{B}\cdot d\mathbf{l}=\epsilon_0\mu_0\frac{d}{dt}\int_S \mathbf{E}\cdot d\mathbf{S}
\end{aligned}
\qquad \textbf{(29-8)}
$$

Note the symmetrical way that **E** and **B** appear in these relations; this feature is a direct consequence of the introduction of the displacement current. Note also that the solution $\mathbf{E}=\mathbf{B}=0$ satisfies these equations. However, there are nonzero solutions as well. These correspond to electromagnetic waves and are the the subject matter of the remainder of this chapter.

Let us describe some general properties of these equations. First, they are *homogeneous* in **E** and **B**, in that if **E** and **B** are two fields that satisfy (29-8) then so do the fields $\alpha\mathbf{E}$ and $\alpha\mathbf{B}$, for α an arbitrary constant. For the special choice $\alpha=0$, we see that consistent with our previous studies, in the absence of charges and currents the electric and magnetic fields may vanish. Second, they are *linear*; that is, if $(\mathbf{E}_1, \mathbf{B}_1)$ and $(\mathbf{E}_2, \mathbf{B}_2)$ separately satisfy (29-8), then so do $(\alpha\mathbf{E}_1+\beta\mathbf{E}_2)$, $(\alpha\mathbf{B}_1+\beta\mathbf{B}_2)$ for arbitrary constants α and β. This linearity property is particularly important and it is related directly to the *superposition principle* for the electric field and the magnetic field. Note that according to this superposition principle *any* linear combination of solutions of (29-8) is also a solution.

To display a particular nonzero solution, let $\mathbf{E}_0$ be an arbitrary, but constant, vector with dimensions of electric field. In terms of $\mathbf{E}_0$ define a Cartesian coordinate system with $\mathbf{E}_0$ along the x-axis, and let **i**, **j**, and **k** form a triad of orthogonal unit vectors along its axes (Figure 29-3a). If h is an arbitrary function of a *single variable*, then:

The electromagnetic fields defined by

$$
\begin{aligned}
\mathbf{E}&=\mathbf{i}E_0h(z-ct)\\
\mathbf{B}&=\mathbf{j}\frac{E_0}{c}h(z-ct)
\end{aligned}
\qquad \textbf{(29-9)}
$$

satisfy the charge-current-free Maxwellian equations in (29-8), *provided that the parameter c is defined by*

$$c=(\mu_0\epsilon_0)^{-1/2} \qquad \textbf{(29-10)}$$

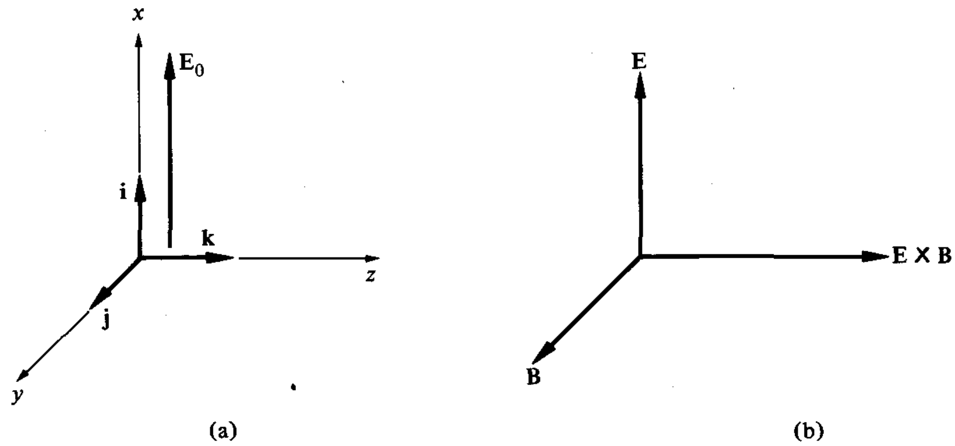


Figure 29-3

In other words, the electromagnetic fields in (29-9) satisfy (29-8) for *any* constant E_0 and *any* function h, provided that the parameter c is defined by (29-10). Inserting the known values $\epsilon_0 = (4\pi \times 9.0 \times 10^9 \text{ N-m}^2/\text{C}^2)^{-1}$ and $\mu_0 = 4\pi \times 10^{-7}$ T-m/A, we find that

$$c = 3.0 \times 10^8 \text{ m/s}$$

which is precisely the speed of light in a vacuum! More generally, as we shall see in Section 29-5, the parameter c is the speed not only of visible light but of *all* electromagnetic waves in a vacuum.

Although the solution in (29-9) has been defined in terms of the coordinate system in Figure 29-3a, the significant features of this solution are actually independent of the system chosen. The following properties deserve particular emphasis:

1. **E** and **B** are mutually perpendicular.
2. **E** and **B** vary spatially only along the direction of $\mathbf{E} \times \mathbf{B}$. If for example, in Figure 29-3b **E** is along the x-axis and **B** is along the y-axis, then **E** and **B** are functions only of z and *not* of x and y.
3. The space-time variation of both **E** and **B** is described in its entirety by the single variable $(z - ct)$, where z is the coordinate along the direction of $\mathbf{E} \times \mathbf{B}$.
4. The ratio of the magnitudes of **E** and **B** is

$$\frac{|\mathbf{E}|}{|\mathbf{B}|} = c = (\mu_0 \epsilon_0)^{-1/2} \tag{29-11}$$

and is thus a *universal constant* whose value can be determined in terms of purely electric and magnetic quantities.

For example, the electromagnetic fields **E** and **B** defined by

$$\mathbf{E} = \mathbf{i} E_0 h(z + ct)$$
$$\mathbf{B} = -\mathbf{j} \frac{E_0}{c} h(z + ct) \tag{29-12}$$

have these properties since the direction of $\mathbf{E}\times\mathbf{B}$ for these is along the *negative* z-axis. Hence the fields in (29-12) also satisfy (29-8) for arbitrary E_0 and h, provided again that c is given by (29-10).

Of particular interest to us is property 3 above. As we discussed in Section 18-4 and, as will be detailed in the next several sections, a space-time variation of this type is *the* distinctive characteristic of wave motion. Hence the electromagnetic fields in (29-9) and (29-12) are known as *electromagnetic waves.*

Unfortunately, it is not easy to verify in detail that (29-9) actually satisfy Maxwell's equations in the form in (29-8). Hence we shall not pursue the matter any further here except to confirm the validity of these solutions for particular choices for S and l.

Example 29-2 Verify that (29-9) satisfy the first two relations in (29-8) for the special case where the closed surface is a cube of side L, as in Figure 29-4.

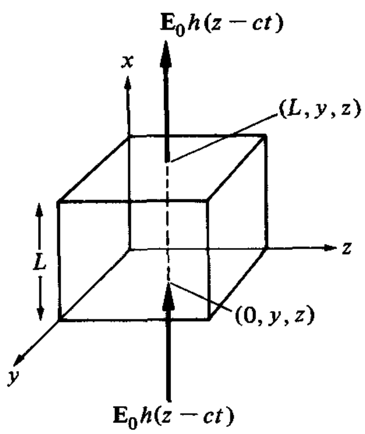


Figure 29-4

Solution Since the direction of $\mathbf{E}$ is along the x-axis, it follows that there is no contribution to the flux integral $\oint_S \mathbf{E}\cdot d\mathbf{S}$ from the front or back face or from either of the side faces. (For each element of area $d\mathbf{S}$ of each of these faces, $\mathbf{E}\cdot d\mathbf{S}=0$.) The flux out of an element of area $dy\,dz$ of the top face with coordinates (L, y, z) is $E_0h(z-ct)dy\,dz$. Correspondingly, the *outward* flux through the element of the bottom face with coordinates $(0, y, z)$ and of area $dy\,dz$ is $-E_0h(z-ct)dy\,dz$, since $\mathbf{E}$ and $d\mathbf{S}$ are antiparallel on this face. Hence

$$\oint_S \mathbf{E}\cdot d\mathbf{S}=\int_0^L dy\int_0^L dz\,E_0h(z-ct)-\int_0^L dy\int_0^L dz\,E_0h(z-ct)=0$$

thus confirming the validity of the first of (29-9) for this special case.

The argument that the magnetic flux out of the cube in Figure 29-4 vanishes follows along the same lines. This time we find an exact cancellation between the flux out of the front and back faces of the cube. The flux out of each of the remaining faces vanishes identically since for each of these $\mathbf{B}\cdot d\mathbf{S}=0$.

Example 29-3 Show that (29-9) satisfies the fourth relation of (29-8) for the special choice where S is a square of side L in the y-z plane (see Figure 29-5).

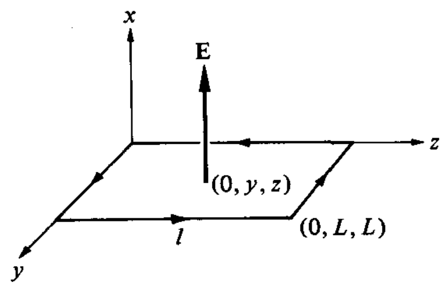


Figure 29-5

Solution Assuming that the positive sense of $d\mathbf{S}$ is along the positive x-axis, the electric flux through the surface is:

$$\int \mathbf{E}\cdot d\mathbf{S} = \int_0^L \int_0^L (\mathbf{i}\,dy\,dz)\cdot(\mathbf{i}E_0 h(z-ct)) = \int_0^L dy \int_0^L dz\, E_0 h(z-ct)$$

$$= E_0 L \int_0^L dz h(z-ct)$$

since $d\mathbf{S} = \mathbf{i}\,dy\,dz$ in this case, and $\mathbf{i}\cdot\mathbf{i} = 1$. Similarly, since $\mathbf{B}$ has a component only along the y-axis, we obtain for the line integral of $\mathbf{B}$ around the perimeter of the square

$$\oint \mathbf{B}\cdot d\mathbf{l} = \int_0^L (\mathbf{j}\,dy)\cdot\left[\mathbf{j}\frac{E_0}{c}h(-ct)\right] + \int_L^0 (\mathbf{j}\,dy)\cdot\left[\mathbf{j}\frac{E_0}{c}h(L-ct)\right]$$

$$= \frac{LE_0}{c}[h(-ct) - h(L-ct)]$$

since $\mathbf{j}\cdot\mathbf{j} = 1$ and since along the left side of the square $z = 0$ while on the right $z = L$.

If we now substitute these formulas into the fourth relation in (29-8) there results

$$\frac{LE_0}{c}[h(-ct) - h(L-ct)] = \epsilon_0\mu_0 \frac{d}{dt} E_0 L \int_0^L dz\, h(z-ct)$$

$$= \epsilon_0\mu_0 E_0 L \int_0^L dz \frac{d}{dt} h(z-ct)$$

$$= \epsilon_0\mu_0 E_0 L \int_0^L dz \left[-c\frac{d}{dz} h(z-ct)\right]$$

$$= -\epsilon_0\mu_0 E_0 L c\, h(z-ct)\Big|_0^L$$

$$= \epsilon_0\mu_0 E_0 L c[h(-ct) - h(L-ct)]$$

where the third equality follows from the identity

$$\frac{d}{dx} f(x+y) = \frac{d}{dy} f(x+y)$$

which is valid for any function f. (See Example 18-5 in this connection.) Finally, then, because of (29-10) this is an identity for all values for the parameters E_0 and L and for any choice of the function h. The validity of (29-9) for the particular surface in Figure 29-5 is thus confirmed.

29-5 The propagation of electromagnetic waves

In order to understand better the physical significance of the solutions in (29-9) and (29-12) let us review some of the important ideas developed in Chapter 18 for wave motion on a string. Consider a string stretched along the x-axis of a certain coordinate system and under a tension T_0 (Figure 18-7). If a disturbance is generated somewhere along the string and propagates, say, along the positive sense of the x-axis, then the displacement $y = y(x, t)$ at time t of an element of the string at the point x may be expressed as

$$y = h(x - ut) \tag{18-5}$$

Correspondingly, for a wave traveling along the negative sense of the x-axis, the displacement is of the form $y = h(x + ut)$. In these formulas, h is an arbitrary function of a single variable and the parameter u characterizes the speed of the wave along the string. (Its value is given by $u = (T_0/\mu)^{1/2}$, where μ is the mass per unit length of the string.) As shown in Figure 18-3, the actual displacements of the constituent particles of the string are perpendicular to the direction of propagation of the wave and thus the wave is *transverse*. By contrast, for a *longitudinal* wave, such as a sound wave in air, the motions of the constituent particles are along the direction of propagation of the wave. Finally, since the motions of the constituent particles in Figure 18-6a are along the y-axis, the wave is also said to be *plane polarized* or *linearly polarized* along this axis. Thus, the wave shown in Figure 18-6b is linearly polarized along the z-axis. As exemplified in Figure 18-6c, not all waves on a string need be plane polarized.

Let us now return to the solution in (29-9) of the charge-current-free Maxwell equations. In light of the preceding discussion, we see that this solution represents a wave of some type traveling at the speed c along the positive sense of the z-axis. We shall refer to it as an *electromagnetic wave.* The "displacement" of the constituents of the wave are the fields **E** and **B** in (29-9) themselves, and since these fields are perpendicular to the direction of propagation of the wave it follows that electromagnetic waves are *transverse.* It is customary to define the direction of polarization of the wave by the direction of the electric vector. Therefore, the particular electromagnetic wave defined by (29-9) is linearly polarized along the x-axis. Similarly, the fields in (29-12) correspond to a transverse wave polarized along the x-axis and traveling along the negative sense of the z-axis.

Very generally, then, the charge-current-free Maxwellian equations in (29-8) admit wave solutions corresponding to transverse electromagnetic waves traveling at the speed c. The electric and magnetic fields **E** and **B** of the wave are perpendicular to each other and to the direction of propagation of the wave, $\mathbf{E} \times \mathbf{B}$. The ratio $|\mathbf{E}|/|\mathbf{B}|$ for all electromagnetic waves traveling through empty space is the constant c, defined by (29-10).

Although the two classes of solutions in (29-9) and (29-12) correspond to

linearly polarized electromagnetic waves, unpolarized solutions of (29-8) also exist. This follows most directly from the fact that linear combinations of solutions of (29-8) are also solutions. For example, the electromagnetic fields given by

$$\mathbf{E} = E_0[\mathbf{i}h_1(z - ct) + \mathbf{j}h_2(z + ct)]$$

$$\mathbf{B} = \frac{E_0}{c}[\mathbf{j}h_1(z - ct) + \mathbf{i}h_2(z + ct)]$$

for h_1 and h_2 arbitrary functions satisfy (29-8) and correspond to an unpolarized electromagnetic wave.

Figure 29-6 shows at $t = 0$ an instantaneous picture of a polarized electromagnetic wave pulse associated with a particular function $h(z)$. The electric field vector of the wave lies in the x-z plane, and the corresponding magnetic field is in the y-z plane. Except for a scale factor, the envelopes of the vectors for these two fields are the same. In the course of time, as the wave travels along the positive sense of the z-axis, the equation for the envelope of the electric vector evolves as $x = h(z - ct)$ and similarly for $\mathbf{B}$. Consistent with our discussion of traveling waves in Chapter 18, we can visualize the propagation of the pulse in Figure 29-6 by imagining the entire pattern to travel along the z-axis at the speed c.

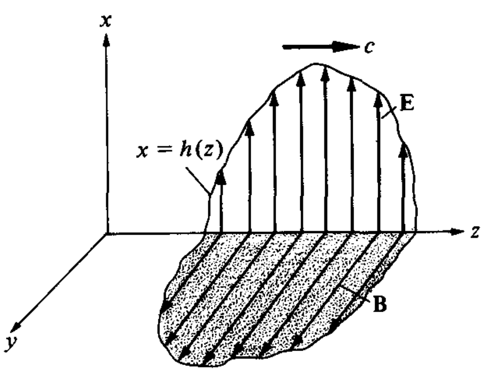


Figure 29-6

Example 29-4 The *index of refraction n* of a substance is defined as the ratio of the speed c of light in a vacuum to its speed v when traveling through that substance:

$$n = \frac{c}{v} \tag{29-13}$$

Calculate the index of refraction of a material characterized by a dielectric constant κ and a relative permeability κ_m.

Solution According to (29-7), in the absence of "true" charges and currents ($q_t = 0$; $\mathbf{j}_t = 0$), the first, second, and third relations of (29-8) still apply, whereas the fourth must be replaced by

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \epsilon_0\mu_0\kappa\kappa_m \frac{d}{dt}\int_S \mathbf{E} \cdot d\mathbf{S}$$

Hence, comparing this relation with (29-8) through (29-10) we see that the speed v of waves in this medium is given by

$$v = [\epsilon_0 \mu_0 \kappa \kappa_m]^{-1/2} = \frac{c}{(\kappa \kappa_m)^{1/2}}$$

since $c = (\epsilon_0 \mu_0)^{-1/2}$. Finally, comparison with (29-13) leads to

$$n = (\kappa \kappa_m)^{1/2} \tag{29-14}$$

29-6 The Poynting vector

As for other types of wave motion, a flow of energy is generally associated with the propagation of an electromagnetic wave. The purpose of this section is to define a quantity known as the *Poynting vector,* in terms of which the flow of electromagnetic energy is conveniently described.

Consider, in Figure 29-7, a region of space through which an electromagnetic wave is propagating, and let S be a closed surface in this region. If **E** and **B** are the fields associated with the wave, then according to (28-15) and the result of Problem 26, Chapter 22, the total energy U inside S is

$$U = \frac{\epsilon_0}{2} \int_V \mathbf{E}^2 \, dV + \frac{1}{2\mu_0} \int_V \mathbf{B}^2 \, dV \tag{29-15}$$

where in both integrals the region of integration is the total volume V inside S. As the wave propagates along, the values of both **E** and **B** inside S will in general change, and associated with this change the electromagnetic energy U inside S will also be modified. It follows that, as the wave propagates along, energy will in general flow across the bounding surface S.

To characterize this energy flow, let us define the *Poynting vector* **N** to represent the magnitude and the direction of the energy flow per unit area per unit time. This means that if, in Figure 29-7, $d\mathbf{S}$ is a vectorial area element of the closed surface S, then the quantity $\mathbf{N} \cdot d\mathbf{S}$ is the outward flow of electromagnetic energy per unit time through $d\mathbf{S}$. In particular, the total flow of energy per unit time out of S is the surface integral

$$\oint_S \mathbf{N} \cdot d\mathbf{S}$$

with the integration extending over the entire closed surface S.

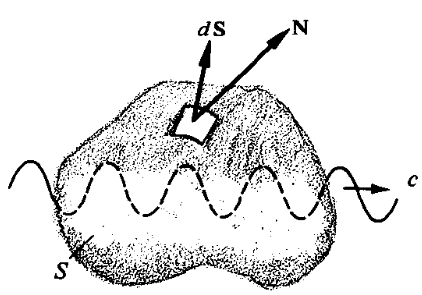


Figure 29-7

Now according to the ideas of energy conservation, if there is an outward flow of energy from S the total electromagnetic energy inside must change. Hence follows the basic relation

$$\frac{d}{dt}U + \oint_S \mathbf{N} \cdot d\mathbf{S} = 0 \tag{29-16}$$

with U defined in (29-15). In physical terms, this states that as an electromagnetic wave propagates along, at each instant the rate at which the energy inside a given surface increases plus the outward flow of energy per unit time through the surface must vanish. Or equivalently, the rate at which the energy increases inside a given surface is numerically equal to the flow of energy per unit time *into* that surface.

By use of the formula for U in (29-15) and the fact that (29-16) is applicable for all closed surfaces S, an explicit formula for the Poynting vector in terms of $\mathbf{E}$ and $\mathbf{B}$ can be derived. The result is

$$\mathbf{N} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B} \tag{29-17}$$

and below we shall confirm the validity of this formula for a special case. Note that the direction of the Poynting vector is along the propagation direction $\mathbf{E} \times \mathbf{B}$ of the electromagnetic wave. Thus the wave defined by (29-9) propagates along the positive z-axis, which is along the direction of $\mathbf{N}$. Similarly, for the wave in (29-12), $\mathbf{E} \times \mathbf{B}$ is along the negative z-axis, consistent with the fact that the associated wave travels along this direction also.

It is interesting to note that momentum is also associated with the flow of electromagnetic energy in a wave. Detailed studies of Maxwell's equations show that if $\mathbf{E}$ and $\mathbf{B}$ are the fields at a given point in space, then the momentum density, or the momentum per unit volume $\mathbf{P}_v$ in that region, is

$$\mathbf{P}_v = \frac{1}{c^2}\mathbf{N} \tag{29-18}$$

with $\mathbf{N}$ given in (29-17). A number of applications of this formula are given in the problems. One important physical prediction of (29-18) is that if an electromagnetic wave falls on matter it exerts mechanical pressure on it. This is known as ***radiation pressure*** and is the basis for a number of suggested proposals for space-propulsion systems that utilize the radiant energy output of the sun.

Example 29-5 Confirm the validity of (29-17) for the solution in (29-9) by using for S the cubical surface in Figure 29-4.

Solution Substituting the given forms for $\mathbf{E}$ and $\mathbf{B}$ into (29-17), we find that

$$\mathbf{N} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B} = \frac{E_0^2}{\mu_0 c} h^2(z - ct)\mathbf{k}$$

with $\mathbf{k} = \mathbf{i} \times \mathbf{j}$ a unit vector along the z-axis. Accordingly, the flow of energy per unit time $\mathbf{N} \cdot d\mathbf{S}$ out of the top, the bottom, and the front and back faces of the cube vanishes. Hence

$$\oint \mathbf{N} \cdot d\mathbf{S} = \frac{E_0^2}{\mu_0 c}\left[\int_0^L \int_0^L dx\, dy\, h^2(L - ct) - \int_0^L \int_0^L dx\, dy\, h^2(-ct)\right]$$
$$= \frac{L^2 E_0^2}{\mu_0 c}[h^2(L - ct) - h^2(-ct)] \tag{29-19}$$

since along the right face $z = L$ and on the left $z = 0$. Similarly, the substitution of the given forms for $\mathbf{E}$ and $\mathbf{B}$ into (29-15) leads to

$$U = \int_0^L \int_0^L \int_0^L dx\, dy\, dz\left[\frac{\epsilon_0}{2} E_0^2 h^2(z - ct) + \frac{1}{2\mu_0}\frac{E_0^2}{c^2} h^2(z - ct)\right]$$
$$= \epsilon_0 E_0^2 L^2 \int_0^L dz\, h^2(z - ct)$$

where the second equality follows from (29-10) and the fact that the integrand is independent of x and y. Making use of the same method as in Example 29-3 we find that

$$\frac{dU}{dt} = \frac{d}{dt}\epsilon_0 E_0^2 L^2 \int_0^L dz\, h^2(z - ct) = \epsilon_0 E_0^2 L^2 \int_0^L dz \frac{d}{dt} h^2(z - ct)$$
$$= -\epsilon_0 E_0^2 L^2 \int_0^L dz\, c\frac{d}{dz} h^2(z - ct) = -\epsilon_0 c E_0^2 L^2 h^2(z - ct)\Big|_0^L$$
$$= -\epsilon_0 c E_0^2 L^2[h^2(L - ct) - h^2(-ct)]$$

The validity of (29-17) for this special case then follows by comparison with (29-19) and the relation $\epsilon_0 c = 1/\mu_0 c$.

29-7 Monochromatic waves

A *monochromatic* electromagnetic wave is analogous to a sinusoidal wave as defined in Chapter 18, and is one for which the function h in (29-9) or (29-12) is either a sine function or a cosine function. The term "monochromatic" refers to the fact that for visible light a sinusoidal wave is one composed of a single color, or frequency.

To be specific, let us make use of a sine function. Comparison with (18-8) shows then that for a monochromatic wave, (29-9) may be expressed as

$$\mathbf{E} = \mathbf{i}E_0 \sin(kz - \omega t)$$
$$\mathbf{B} = \mathbf{j}\frac{E_0}{c}\sin(kz - \omega t) \tag{29-20}$$

where k and ω are two parameters, known respectively as the *wave number* and the *angular frequency* of the wave. Since the space-time variation of an electromagnetic wave can only be in terms of the variable $(z - ct)$, or

$(z + ct)$, it follows that ω and k must be related by (compare (18-9))

$$\omega = kc \tag{29-21}$$

Hence only one of ω or k is required to specify the wave. Figure 29-8 shows an instantaneous picture of the linearly polarized electromagnetic wave defined by the fields in (29-20). The temporal evolution of the wave may be visualized by imagining the given pattern to be moving along the positive sense of the z-axis at the speed c.

The *wavelength* λ of a monochromatic wave is defined as the distance between two neighboring maxima, or minima, of the wave at a fixed time t; see Figure 29-8. As in the derivation of (18-10), it follows that λ and k are related by

$$\lambda = \frac{2\pi}{k} \tag{29-22}$$

and thus a knowledge of either one is tantamount to a knowledge of the other.

A fourth parameter of interest in connection with a monochromatic wave is its frequency ν. This is defined so that $1/\nu$ is the period of the wave, and thus $1/\nu$ represents the minimum positive time interval after which the wave repeats its form at each point of space. Substitution of this definition into (29-20) leads to the identification $\nu = \omega/2\pi$, and combining this with (29-22) enables us to express (29-21) equivalently as

$$\lambda\nu = c \tag{29-23}$$

To summarize, then, a monochromatic electromagnetic wave, such as the one described by (29-20) is specified by two independent parameters. One of these is its *amplitude* E_0, which, as we shall see in the next section, determines the amount of energy transported by the wave. The second is usually selected to be its wavelength λ or its frequency ν. These are related by (29-23).

Figure 29-9 shows diagrammatically the wavelength intervals and the associated frequency ranges for all known types of electromagnetic radiation. The totality of these radiations constitute the *electromagnetic spec-*

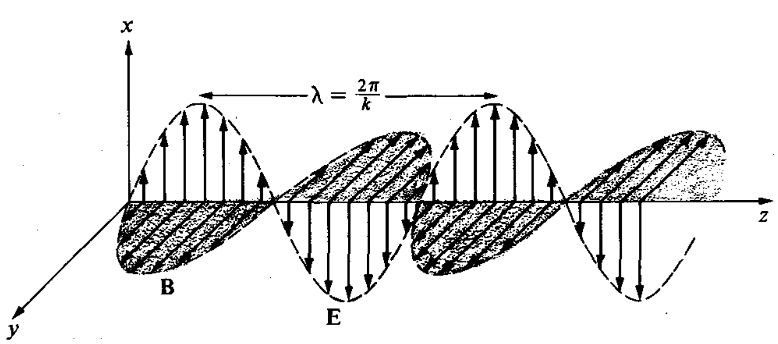


Figure 29-8

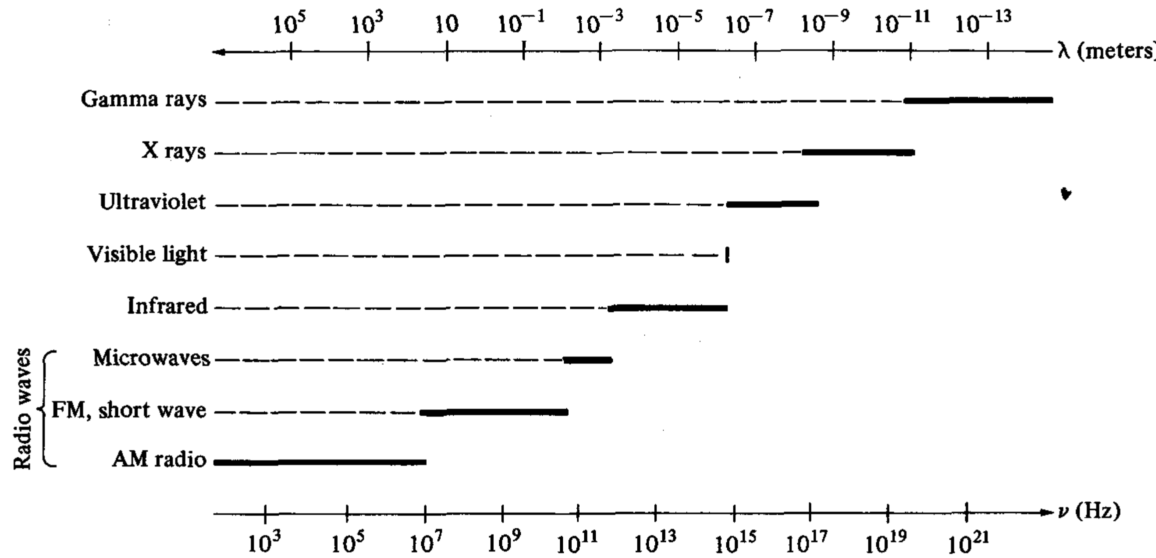


Figure 29-9

trum. At the long-wavelength end of this spectrum are *radio waves*, which are characterized by wavelengths typically of the order of 10 meters and longer, while at the other extreme we find *gamma rays*, with wavelengths of the order of 10^{-11} meter and shorter. Between these two extremes the spectrum is composed of *X rays*, *ultraviolet light*, *visible light*, and *infrared radiation*. Note that in the figure λ increases towards the left, while ν increases towards the right. The unit of frequency is the hertz (Hz), which has been defined in Chapter 18 as 1 cycle per second. Thus since $c = 3.0 \times 10^8$ m/s it follows from (29-23) that the frequency associated with a radio wave of wavelength $\lambda = 20$ meters is

$$\nu = \frac{c}{\lambda} = \frac{3.0 \times 10^8 \text{ m/s}}{20 \text{ m}} = 1.5 \times 10^7 \text{ Hz}$$

Of particular interest to us is the portion of the electromagnetic spectrum that consists of visible light. The wavelengths of electromagnetic waves to which the human eye is sensitive lie along the very restricted interval

$$4.5 \times 10^{-7} \text{ m} \leqslant \lambda \leqslant 6.8 \times 10^{-7} \text{ m}$$

The shortest wavelengths of visible light are associated with violet light and for increasing values of λ, the color changes continuously from blue through green, yellow, orange, and finally to red light, which is associated with a wavelength of approximately 6.8×10^{-7} meter. The frequencies associated with visible light are found by use of (29-23) to lie along the interval

$$6.7 \times 10^{14} \text{ Hz} \geqslant \nu \geqslant 4.4 \times 10^{14} \text{ Hz}$$

Figure 29-10 shows schematically a plot of the energy emitted by our sun as a function of wavelength. It is interesting to note that this curve shows a peak at wavelengths that correspond to visible light. That is, the spectrum of

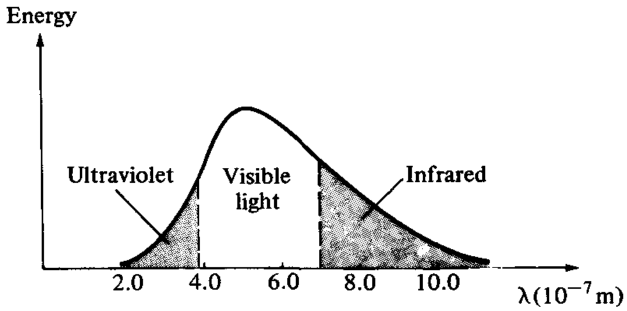


Figure 29-10

the radiation emitted by our sun peaks at precisely those wavelengths where the eye is most sensitive. By a happy coincidence (!) the earth's atmosphere also happens to be transparent for these same wavelengths. Almost all of the ultraviolet light emitted by the sun and much of that radiated in the infrared is absorbed by the earth's atmosphere. Several decades ago it was found that the atmosphere is also transparent to radiation of wavelengths in the range from 10^{-2} to 10^2 meters. This very important discovery ushered in the era of *radio astronomy*. Fortunately, the atmosphere absorbs radiation of all wavelengths except for radio waves and visible light. Life, as we know it on earth, would not be possible if the atmosphere were transparent to ultraviolet light or X rays.

Example 29-6 Consider yellow light of wavelength 5.7×10^{-7} meter. Calculate:

(a) Its frequency.
(b) Its period.
(c) The wave number associated with it.

Solution

(a) Solving (29-23) for ν and substituting the given values for c and λ, we find that

$$\nu = \frac{c}{\lambda} = \frac{3.0 \times 10^8 \text{ m/s}}{5.7 \times 10^{-7} \text{ m}} = 5.3 \times 10^{14} \text{ Hz}$$

(b) The period T is the reciprocal of ν; hence

$$T = \frac{1}{\nu} = \frac{1}{5.3 \times 10^{14} \text{ Hz}} = 1.9 \times 10^{-15} \text{ s}$$

(c) Substituting the given value for λ into (29-22), we obtain

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{5.7 \times 10^{-7} \text{ m}} = 1.1 \times 10^7/\text{m}$$

Example 29-7 An electromagnetic wave of wavelength λ_0 and of frequency $\nu_0 = c/\lambda_0$ originally traveling in free space enters a block of matter characterized by a dielectric constant κ. Assuming that $\kappa_m \approx 1$:

(a) What is the speed of the wave in the material?

(b) Explain physically why the frequency of the wave must be the same inside and outside of the material.
(c) Find the wavelength of the wave inside the material.

Solution
(a) According to (29-14), the index of refraction n of the material is $\sqrt{\kappa}$ since we have assumed that $\kappa_m \approx 1$. Hence substitution into (29-13) yields for the velocity v of the wave in the medium

$$v = \frac{c}{\sqrt{\kappa}}$$

(b) Since the incident electromagnetic fields oscillate at the frequency ν_0, the induced dipole moment inside the material will also oscillate at this same frequency. It follows that the total electromagnetic field inside the material—which is the sum of the external fields plus those produced by the dipoles—will also oscillate with the same frequency ν_0. Hence the frequency of the wave inside and outside of this material is the same.
(c) The relation, $\nu\lambda = c$ in (29-23) is very generally valid for any wave. Thus, since the velocity of the wave in the medium is $c/\sqrt{\kappa}$, while its frequency according to (b) is ν_0, it follows that the wavelength λ_m inside the medium is

$$\lambda_m = \frac{v}{\nu_0} = \frac{c}{\nu_0\sqrt{\kappa}} = \frac{\lambda_0}{\sqrt{\kappa}}$$

where in the final equality we have used the fact that the free-space wavelength $\lambda_0 = c/\nu_0$. Generally $\kappa > 1$, so the wavelength λ_m inside the medium is less than the wavelength λ_0 in free space. Equivalently, this result may be expressed in terms of the index of refraction

$$\lambda_m = \frac{\lambda_0}{n} \tag{29-24}$$

since $n = \sqrt{\kappa}$ in this case.
For example, if light of wavelength 5.5×10^{-7} meter in free space enters a medium of refractive index $n = 1.5$, its wavelength in the medium, according to (29-24), is

$$\lambda_m = \frac{\lambda_0}{n} = \frac{5.5 \times 10^{-7}\text{ m}}{1.5} = 3.7 \times 10^{-7}\text{ m}$$

and this no longer corresponds to visible light. The velocity of light v in this medium is

$$v = \frac{c}{n} = \frac{3.0 \times 10^8\text{ m/s}}{1.5} = 2.0 \times 10^8\text{ m/s}$$

29-8 Energy propagation in a monochromatic wave

According to Figure 29-9, the frequency of visible light is of the order of 10^{15} Hz. This means that both the electric and the magnetic vectors of the wave reverse themselves about 10^{15} times each second. Since this is very large compared to the frequencies associated with the motion of macro-

scopic matter, including notably those associated with our sensory organs, it follows that when light enters our eye what we observe is radiation averaged over many cycles. Hence for purposes of discussing the energy flow associated with visible light, as well as certain other parts of the electromagnetic spectrum, only the time average of the Poynting vector is of physical interest. The purpose of this section is to show that this time average $\bar{N}$ of the magnitude of the Poynting vector for a monochromatic wave is

$$\bar{N} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\, E_0^2 \qquad (29\text{-}25)$$

where E_0 is the amplitude of the wave.

To this end consider the monochromatic wave characterized by the amplitude E_0 and wave number k in (29-20). According to (29-21) through (29-23), this may be expressed equivalently as

$$\mathbf{E} = \mathbf{i}E_0 \sin\left[\frac{2\pi}{\lambda}(z - ct)\right]$$

$$\mathbf{B} = \mathbf{j}\frac{E_0}{c} \sin\left[\frac{2\pi}{\lambda}(z - ct)\right]$$

Substitution into (29-17) shows that the direction of energy flow is along the z-axis and that the associated magnitude N of the Poynting vector is

$$N = \left|\frac{1}{\mu_0}\mathbf{E}\times\mathbf{B}\right| = \frac{E_0^2}{\mu_0 c}\sin^2\left[\frac{2\pi}{\lambda}(z - ct)\right]$$

Note that, in general, N varies with both position and time.

To calculate the time average $\bar{N}$ of the Poynting vector $\mathbf{N}$, let us make use of the fact that the average value $\bar{f}$ of a function $f(t)$ over a fixed interval $a < t < b$ is defined by

$$\bar{f} = \frac{1}{(b-a)}\int_a^b f(t)\, dt$$

Applying this definition to the Poynting vector, we find that the average value of N over a period $T = 1/\nu$ is

$$\bar{N} = \nu\int_0^{1/\nu} N\, dt = \frac{E_0^2}{\mu_0 c}\,\nu\int_0^{1/\nu} \sin^2\left[\frac{2\pi}{\lambda}(z - ct)\right] dt \qquad (29\text{-}26)$$

In the problems it will be established that the value of this integral is $1/2\nu$ and is thus independent of spatial variable z. Hence, since $\epsilon_0\mu_0 c^2 = 1$, the relation in (29-25) follows.

It is also shown in the problems that for a wave composed of more than one frequency, the right-hand side of (29-25) must be replaced by a sum of terms, one for each frequency. In this connection it should be noted that (29-25) is independent of λ and is thus the same for all frequencies. Hence, the ratio of the energies associated with two waves of different frequencies

is independent of the frequencies involved; it depends only on the ratio of the squares of the amplitudes of the two waves.

It should be emphasized that (29-25) applies only to physical situations for which the observer is so far from the source that (29-20) is applicable "everywhere." More realistically, as will be shown in the following section, $\bar{N}$ varies inversely with the square of the distance from the source. For example, since the planet Jupiter is about 5.2 times as far away from the sun as is the earth, it follows that the solar energy incident per unit time on a unit area of Jupiter is $(1/5.2)^2 = 0.037$ of that falling on the earth. In applying (29-25) to practical situations this feature must be kept in mind.

Example 29-8 A 1-mW helium-neon laser sends out a continuous beam of red light. Assuming that the beam has a 0.2-cm² cross section and does not appreciably diverge, calculate:

(a) The average energy flux $\bar{N}$.
(b) The amplitude E_0 of the associated wave.
(c) The strength of the magnetic induction B_0 of the wave.

Solution

(a) Since the output of the laser is 1 mW and since the area of the beam is $0.2\ \text{cm}^2 = 2.0 \times 10^{-5}\ \text{m}^2$, the flux or the energy per unit area per unit time, $\bar{N}$, is

$$\bar{N} = \frac{1.0 \times 10^{-3}\ \text{W}}{2.0 \times 10^{-5}\ \text{m}^2} = 50\ \text{W/m}^2$$

(b) Solving (29-25) for E_0, and substituting the above value for $\bar{N}$, we obtain

$$E_0 = \left[2\bar{N}\sqrt{\frac{\mu_0}{\epsilon_0}}\right]^{1/2} = \left[2 \times 50\ \text{W/m}^2 \times \left(\frac{4\pi \times 10^{-7}\ \text{T-m/A}}{8.9 \times 10^{-12}\ \text{C}^2/\text{N-m}^2}\right)^{1/2}\right]^{1/2}$$
$$= 1.9 \times 10^2\ \text{V/m}$$

(c) Since for an electromagnetic wave the amplitude B_0 of the **B**-field is E_0/c, it follows that

$$B_0 = \frac{E_0}{c} = \frac{1.9 \times 10^2\ \text{V/m}}{3 \times 10^8\ \text{m/s}} = 6.3 \times 10^{-7}\ \text{T}$$

†29-9 The sources of electromagnetic waves

Having demonstrated that the charge-current-free Maxwell equations admit solutions corresponding to electromagnetic waves, in this section we briefly consider the problem of how such waves are generated.

In our previous studies of Maxwell's equations with the displacement current term omitted, we have seen that the electromagnetic fields associated with various charge-current distributions go to zero very rapidly far away from these sources. Specifically, if r is the separation distance between the observation point and the source, then both **E** and **B** fall off as

$1/r^2$ as r becomes very large. In turn, this means that the Poynting vector **N** in (29-17) falls off as $1/r^4$ and, as will be seen below, no electromagnetic waves are associated with such fields. However, if the displacement current term is included in Maxwell's equations, then under certain circumstances there are solutions of these equations that at large separation distances from the source fall off only as $1/r$. For these solutions—which are known as the *radiation fields*—the Poynting vector varies as $1/r^2$, and associated with these fields there is an unambiguous flow of energy.

Consider, in Figure 29-11, a charge-current source located at the origin of a certain coordinate system. Assuming that for very large distances r from this source both **E** and **B** vary as $1/r$, it follows from (29-17) that

$$\mathbf{N} \cong \mathbf{N}_0 \frac{1}{r^2} \qquad (r \to \infty) \tag{29-27}$$

with $\mathbf{N}_0$ a vector independent of r. In general, however, $\mathbf{N}_0$ will vary with the angles (θ, ϕ) as defined in the figure. Now in terms of these angles, an area element $d\mathbf{S}$ of the sphere is

$$d\mathbf{S} = \mathbf{e}r^2 \sin\theta \, d\theta \, d\phi$$

with **e** a unit vector along the radial direction. Hence the electromagnetic energy per unit time that passes through $d\mathbf{S}$ is (for $r \to \infty$)

$$\begin{aligned} \mathbf{N} \cdot d\mathbf{S} &= \left(\mathbf{N}_0 \frac{1}{r^2}\right) \cdot (\mathbf{e}r^2 \sin\theta \, d\theta \, d\phi) \\ &= (\mathbf{N}_0 \cdot \mathbf{e}) \sin\theta \, d\theta \, d\phi \end{aligned} \tag{29-28}$$

and this is independent of r. Thus we have the result:

Regardless of the separation distance between the observer and the source, provided **E** *and* **B** *both vary as* $1/r$ *for large* r*, electromagnetic energy will be radiated by the source unless* $|\mathbf{N}_0| \equiv 0$.

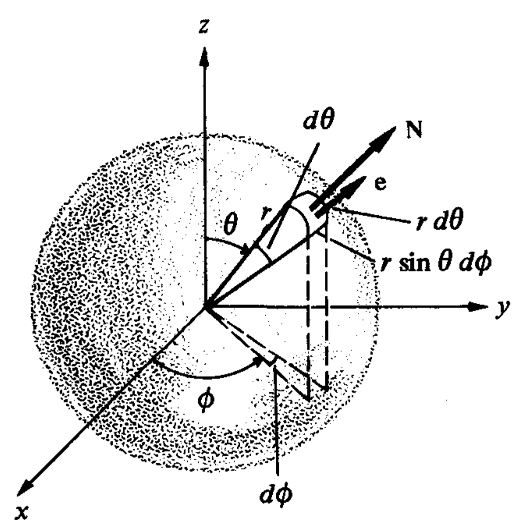


Figure 29-11

The *total* rate dW/dt at which energy is radiated by the source may be obtained by integrating (29-28) over the entire sphere. The result is

$$\frac{dW}{dt} = \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta \, d\theta \, \mathbf{N}_0 \cdot \mathbf{e} \qquad \text{(29-29)}$$

To illustrate the types of charge-current distributions that radiate electromagnetic energy, consider in Figure 29-12 a particle of charge q which has an acceleration $\mathbf{a}$. For distances r very far from the particle, the electric and magnetic fields are found to be

$$\begin{aligned} \mathbf{E} &= \frac{q}{4\pi\epsilon_0 c^2 r^3} \mathbf{r} \times (\mathbf{r} \times \mathbf{a}) \\ \mathbf{B} &= \frac{q}{4\pi\epsilon_0 c^3 r^2} \mathbf{a} \times \mathbf{r} \end{aligned} \qquad \text{(29-30)}$$

provided that the speed of the particle is small compared to c. Substituting these formulas into (29-27), we find that

$$\mathbf{N}_0 = \mathbf{e} \frac{q^2}{16\pi^2 \epsilon_0} \frac{a^2}{c^3} \sin^2\theta$$

with θ as defined in the figure. The energy radiated by the particle vanishes along the direction of the acceleration, and is a maximum in a direction perpendicular to $\mathbf{a}$. The rate at which the particle radiates energy in all directions is found by substitution into (29-29) to be

$$\frac{dW}{dt} = \frac{q^2 a^2}{6\pi\epsilon_0 c^3} \qquad \text{(29-31)}$$

Note the important feature that if $a = 0$, then no energy is radiated. In other words, in order for a charged particle to radiate energy it is *necessary that it accelerate.*

Correspondingly, in order for a macroscopic current distribution to radiate electromagnetic energy, it is necessary that the current vary in time. For example, if a current $i = i_0 \cos\omega t$ flows around a current loop of area A,

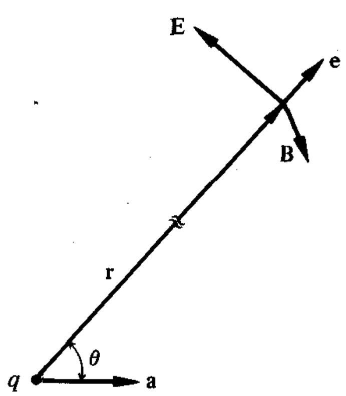


Figure 29-12

then the rate at which energy is radiated is found to be

$$\frac{dW}{dt} = \frac{1}{3}\sqrt{\frac{\mu_0}{\epsilon_0}}(i_0 A)^2 \left(\frac{\omega}{c}\right)^4$$

Example 29-9 At what rate is energy being radiated by a proton in a cyclotron at an instant when it is undergoing a centripetal acceleration of 2.0×10^{15} m/s^2?

Solution Substituting the given data into (29-31), we obtain

$$\begin{aligned}\frac{dW}{dt} &= \frac{2}{3}\frac{q^2}{4\pi\epsilon_0}\frac{a^2}{c^3} \\ &= \frac{(0.67) \times (1.6 \times 10^{-19}\ \text{C})^2 \times (9.0 \times 10^9\ \text{N-m}^2/\text{C}^2) \times (2.0 \times 10^{15}\ \text{m/s}^2)^2}{(3.0 \times 10^8\ \text{m/s})^3} \\ &= 2.3 \times 10^{-23}\ \text{W}\end{aligned}$$

†29-10 The displacement current

The purpose of this section is to consider in a quantitative way the difficulty associated with Ampère's law and to show how this is resolved by the introduction of the displacement current.

Imagine, in Figure 29-13, applying Ampère's law to a sequence of closed curves $l_1, l_2, \ldots$, whose lengths, as in the figure, approach zero. In this limit, suppose that the sequence of surfaces $S_1, S_2, \ldots$ bounded by these curves approaches a *finite* closed surface S. Since the magnetic induction **B** is generally finite, it follows that the sequence of numbers

$$\oint_{l_1} \mathbf{B} \cdot d\mathbf{l}_1, \quad \oint_{l_2} \mathbf{B} \cdot d\mathbf{l}_2, \quad \ldots$$

tends to zero because the sequence of path lengths does. Accordingly, in this limit the left-hand side of (29-4) vanishes, and since the limiting surface S is closed, we find

$$\oint_S \mathbf{j} \cdot d\mathbf{S} = 0 \qquad \textbf{(29-32)}$$

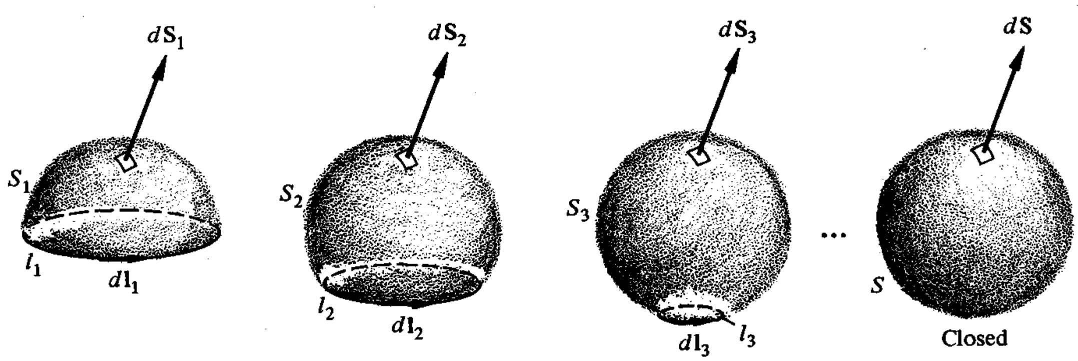


Figure 29-13

But the integral in this relation represents the net flow of current outward through the *closed* surface S. Hence, we conclude that Ampère's law implies of necessity that *the net current flow out of every closed surface must vanish.*

But this cannot always be true! Consider again, in Figure 29-2, a capacitor that is being charged up by a current i, and let S be a surface that encloses the positively charged plate in its entirety. The rate at which the charge on the plate increases is dq/dt, so the current flowing *out* of the closed surface is $-dq/dt$. Hence, for this case,

$$\oint_S \mathbf{j} \cdot d\mathbf{S} = -\frac{dq}{dt}$$

and this is in direct contradiction with the implication of Ampère's law in (29-32).

Based on arguments of this type, Maxwell reasoned that Ampère's law in the form in (29-4) could not be correct. According to the ideas of charge conservation, the outward flow of current from a closed surface is zero only if the rate of change of charge inside that surface vanishes. More generally, the correct form of (29-32) must be

$$\oint_S \mathbf{j} \cdot d\mathbf{S} = -\frac{d}{dt} q(S) \tag{29-33}$$

with $q(S)$ the total charge inside S. Hence, to be consistent with experiment it is necessary to modify Ampère's law so that its limiting form for a closed surface does *not* lead to the generally incorrect (29-32), but rather to the correct version in (29-33).

Let us now confirm that the modified form of Ampère's law in (29-5) does not suffer from this difficulty. To this end, let us apply (29-5) to the sequence of surfaces in Figure 29-13. As in the above derivation of (29-32), the left-hand side of (29-5) vanishes in this limit so that this time there results

$$\oint_S \mathbf{j} \cdot d\mathbf{S} = -\epsilon_0 \frac{d}{dt} \oint_S \mathbf{E} \cdot d\mathbf{S}$$

where S is an arbitrary closed surface. Substituting on the right-hand side by use of Gauss' law, (29-1), we find that

$$\begin{aligned} \oint_S \mathbf{j} \cdot d\mathbf{S} &= -\epsilon_0 \frac{d}{dt} \oint_S \mathbf{E} \cdot d\mathbf{S} = -\epsilon_0 \frac{d}{dt} \left[\frac{1}{\epsilon_0} q(S) \right] \\ &= -\frac{d}{dt} q(S) \end{aligned}$$

and this, at least, is consistent with (29-33) and the law of conservation of electric charge. The detailed justification for the validity of (29-5) is founded, of course, on a much broader experimental basis than is implied by this calculation.

29-11 Summary of important formulas

Maxwell's equations are

$$\oint_S \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} q(S) \tag{29-1}$$

$$\oint_S \mathbf{B} \cdot d\mathbf{S} = 0 \tag{29-2}$$

$$\oint_l \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S} \tag{29-3}$$

$$\oint_l \mathbf{B} \cdot d\mathbf{l} = \mu_0 \int_S \mathbf{j} \cdot d\mathbf{S} + \epsilon_0 \mu_0 \frac{d}{dt} \int_S \mathbf{E} \cdot d\mathbf{S} \tag{29-5}$$

where the symbols S and l are as defined in Figure 29-1. The quantity $q(S)$ represents the *total* charge inside the closed surface S and $\mathbf{j}$ is the current density.

In regions of space far from charges and currents, Maxwell's equations admit wave solutions of the form

$$\begin{aligned} \mathbf{E} &= \mathbf{i} E_0 h(z - ct) \\ \mathbf{B} &= \mathbf{j} \frac{E_0}{c} h(z - ct) \end{aligned} \tag{29-9}$$

with h an arbitrary function and c defined by

$$c = (\epsilon_0 \mu_0)^{-1/2} \tag{29-10}$$

Associated with these waves there is an energy flow given by the *Poynting vector*

$$\mathbf{N} = \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B} \tag{29-17}$$

which represents the rate of energy flow per unit area along the direction of propagation of the wave. For a monochromatic wave, the time average $\bar{N}$ of the magnitude of the Poynting vector is independent of wavelength and is

$$\bar{N} = \frac{1}{2} \sqrt{\frac{\epsilon_0}{\mu_0}} E_0^2 \tag{29-25}$$

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** displacement current; **(b)** Maxwell's equations; **(c)** index of refraction; and **(d)** electromagnetic spectrum.
2. By starting with Maxwell's equations, state what restrictions you must impose on the various physical quantities $\mathbf{E}$, $\mathbf{B}$, $\mathbf{j}$, and ρ so that the equations of electrostatics result.

3. Repeat Question 2, but this time for the equations governing the relation between time-independent currents and their associated magnetic fields.
4. Explain in physical terms why the displacement current term in (29-5) does not appreciably alter the **B**-field associated with a slowly varying current distribution.
5. Under what circumstances is Kirchhoff's first rule—which states that the algebraic sum of the currents approaching a junction must vanish—consistent with Maxwell's equations?
6. Does the argument in Section 29-10 yield a unique value for the displacement current? Give an example of a different term that could be added to Ampère's law so that (29-33) is still satisfied.
7. In Chapter 18, we defined wave motion as "the propagation of motion or of energy through a medium without an associated propagation of matter." Describe in what respects this definition applies to an electromagnetic wave and in what respects it does not.
8. In our study of dielectrics it was noted that the dielectric constant for ordinary water is 80. According to (29-14), it follows that the velocity v of visible light in water would be

$$v = \frac{c}{\sqrt{\kappa}} \cong \frac{c}{9}$$

Explain in physical terms why it is that the measured velocity is actually about $(3/4)c$. (*Hint*: Think of the physical meaning of dielectric constant for a time-dependent electric field.)
9. State a typical value for the wavelength of each of the following types of radiation: **(a)** visible light; **(b)** X rays; **(c)** ultraviolet light; and **(d)** infrared radiation.
10. State a typical value for the frequency associated with each of the following: **(a)** visible light; **(b)** gamma rays; **(c)** radio waves; and **(d)** ultraviolet light.
11. What is the frequency of an AM radio station in your community? What is the wavelength of the radiation associated with this frequency?
12. Consider an electromagnetic wave characterized by the electric field

$$\mathbf{E} = \mathbf{j}E_0 h(x - ct)$$

What is the direction of propagation of the wave? Along what direction does the associated **B**-field vector point?
13. What is the direction of the electric field associated with an electromagnetic wave if the **B**-field of the wave has the form

$$\mathbf{B} = -\mathbf{k}\frac{E_0}{c} h(y + ct)$$

14. Consider an electromagnetic wave characterized by the electric field

$$\mathbf{E} = \mathbf{i}E_0 \cos(kz - \omega t + \alpha)$$

where E_0, k, ω, and α are fixed constants.
   (a) What is the amplitude of this wave?
   (b) Why is the speed of propagation ω/k?
   (c) Need the velocity ω/k be equal to c? Explain.
15. Does an electromagnetic wave always propagate along a straight line? Explain.
16. According to classical ideas, a hydrogen atom consists of a proton and an electron orbiting about it. Why does the electron have to accelerate in this orbit? Show that, in light of (29-31), this atom is unstable; that is, show that the electron would have to radiate energy and thus gradually spiral into the proton. (*Note*: This was one of several incorrect deduc-

tions of the classical laws, which argued against the validity of classical physics, and which led eventually to the development of quantum mechanics.)

17. An electromagnetic wave propagates at the speed c in a given coordinate system. A second observer who is traveling at a speed $u = c/2$ along the propagation direction also observes that the velocity of this wave is c. Can you account for this "paradox" in terms of classical ideas? (*Note*: It was this type of failure of the ideas of the Galilean transformation that helped to bring forth the theory of relativity.)

18. An electron is traveling at a uniform speed $u \sim c$ through a medium of index of refraction n. If u is greater than the speed of light c/n in this medium it is observed that the electron emits electromagnetic radiation known as *Cerenkov radiation*. Give a physical explanation for the existence of this radiation.

19. An electron traveling at a speed $u \sim c$ traverses the interior of the eyeball of a human. In light of Question 18, explain why he will observe a light flash, even though his eyes are closed.

20. Two electromagnetic waves are traveling in mutually perpendicular directions in free speace. If the waves overlap in some region, do you expect them to modify each other in any way? Explain.

21. An electromagnetic wave enters a block of matter of dielectric constant κ and of index of refraction $n = \sqrt{\kappa}$. Explain in microscopic terms why the wave is modified inside the block. Can you account for the fact that the speed of the wave is less than that in free space?

22. Describe the circumstances under which a charged particle will radiate electromagnetic energy. What must happen to the mechanical energy of the particle? To its velocity? To its momentum?

23. In light of Question 22, discuss why do you think that momentum must be associated with an electromagnetic wave.

24. What is the direction of the momentum in an electromagnetic wave traveling radially outward from the sun? Will you experience a force when sunlight shines on you? Explain.

PROBLEMS

1. The sensitivity of a healthy human eye is greatest for green-yellow light of wavelength 5.5×10^{-7} meter. What is the frequency of this light? What is the associated wave number?

2. What is the wavelength of a beam of X rays that are characterized by a frequency of 2.5×10^{19} Hz? What is the angular frequency of these X rays?

3. A monochromatic electromagnetic wave is characterized by a wavelength of 6.0×10^{-7} meter in free space.
   (a) What is the frequency of this wave?
   (b) What is the angular frequency of the wave?
   (c) What is the wave number?

4. Suppose that the wave in Problem 3 enters a region of space containing a nonmagnetic material of dielectric constant $\kappa = 4$.
   (a) What is the index of refraction of this medium?

(b) What is the velocity of the wave in this material?
(c) What is the wavelength inside this matter?

5. A monochromatic wave has associated with it an average energy flux of $12\ W/m^2$.
   (a) What is the strength of the electric field of the wave?
   (b) What is the strength of the magnetic induction?
   (c) Can the frequency of this wave be determined from the given data? Explain.
6. Calculate the strength of the electric and magnetic fields in the radiation reaching us from the sun, assuming for the "solar constant" (the average amount of energy per unit time incident normally on unit area just outside the earth's atmosphere) a value of $1.37 \times 10^3\ W/m^2$ and assuming the sun to be a monochromatic source. Neglect atmospheric absorption.
7. Using the data in Problem 6, calculate the energy flux incident on 1 square meter of Saturn and of Pluto, assuming that these planets are, respectively, 9.2 and 40 times as far away from the sun as is the earth.
8. A monochromatic point source radiates energy in all directions at the average rate of 250 watts.
   (a) What is the average flux of energy at a point 2 meters from the source?
   (b) What is the amplitude of the electric field at this point?
   (c) What is the magnitude of the magnetic induction at a point 5 meters from the source?
9. A monochromatic source emits visible light with a wavelength (in air) of 5.5×10^{-7} meter. What is the wavelength of this light in water? (Assume an index refraction for water of 4/3.) What wavelength would an underwater observer see?
10. Show that for an electromagnetic wave, the energy density $\epsilon_0 E^2/2$ associated with the electric field is the same as the magnetic energy density $B^2/2\mu_0$.
11. Suppose the parallel-plate capacitor in an R-C circuit has plates of area A and separation distance d. If $\mathscr{E}$ is the emf of the battery and current first starts to flow at $t = 0$:
    (a) Show that the electric field E between the plates is
    $$E = \frac{\mathscr{E}}{d}[1 - e^{-t/RC}]$$
    with $C = \epsilon_0 A/d$ the capacitance.
    (b) Calculate the value of the quantity $\epsilon_0 dE/dt$ associated with this field.

*12. Making use of the results of Problem 11 show explicitly that the quantity
    $$\oint_S \left[\mathbf{j} + \epsilon_0 \frac{d\mathbf{E}}{dt}\right] \cdot d\mathbf{S}$$
    vanishes for the closed surface S in Figure 29-2.

13. Show that the electric field $\mathbf{E}$ of an electromagnetic wave propagating along the z-axis satisfies the wave equation
    $$\frac{\partial^2}{\partial z^2}\mathbf{E} = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{E}$$
    Will the $\mathbf{B}$-field of an electromagnetic field satisfy this relation? Prove your answer.
14. Consider, in Figure 29-14, a segment

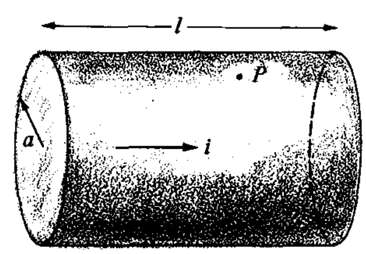


Figure 29-14

of length l of a very long wire of resistance R and radius a, through which there flows a uniform current i.

(a) What is the magnetic induction $\mathbf{B}$ at a point P on the surface of the wire?

(b) Show that the magnitude of the electric field at point P is Ri/l.

(c) What is the magnitude and the direction of the Poynting vector at this point?

(d) Show thus that the rate at which energy flows out of this surface is $-Ri^2$ and give a physical interpretation of this result.

15. Establish the relation

$$\int_0^{1/\nu} \sin^2\left[\frac{2\pi}{\lambda}(z-ct)\right] dt = \frac{1}{2\nu}$$

and thus show the equivalence of (29-25) and (29-26).

16. Consider two monochromatic waves of frequencies ν_1 and ν_2 and of amplitudes E_{01} and E_{02} traveling along the same direction. Show that the time average of the Poynting vector $\bar{N}$ is

$$\bar{N} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}(E_{01}^2 + E_{02}^2)$$

17. Starting with (29-18), show that the time average of the magnitude of the momentum density $\mathbf{P}_v$ of a monochromatic wave is

$$\bar{P}_v = \frac{\epsilon_0 E_0^2}{2c}$$

where E_0 is the amplitude of the wave.

18. Consider, in Figure 29-15, a monochromatic electromagnetic wave of amplitude E_0 traveling perpendicular to a material surface of area A.

(a) Using the result of Problem 17, calculate the time average of the momentum inside the volume $Ac\Delta t$ in the figure.

(b) Assuming that all of this radia-

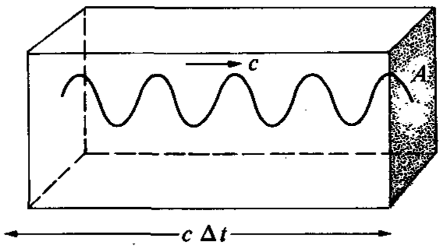


Figure 29-15

tion is absorbed at the surface during the time interval Δt, show that the "radiation pressure" P_R on the surface is

$$P_R = \frac{1}{2}\epsilon_0 E_0^2 = \frac{1}{c}\bar{N}$$

(*Hint*: What is the change in momentum per unit time of the radiation?)

(c) If the radiation is reflected by the surface, show that the radiation pressure is

$$P_R = \frac{2}{c}\bar{N}$$

19. How large must the "reflecting sail" on a spaceship of total mass 5.0×10^3 kg be so that it will accelerate radially outward from the sun at $1.0\ \text{m/s}^2$. Make use of the result of Problem 18 and assume an incident solar flux of $1.0\ \text{kW/m}^2$. Neglect the gravitational attraction of the sun.

20. Using the value for the solar constant in Problem 6 and the result of Problem 18, calculate the force on the earth due to radiation pressure from the sun. Assume the earth to be a perfectly absorbing flat disk of radius 6.4×10^3 km with its plane perpendicular to the sun's rays. Compare this with the gravitational attraction between the earth and the sun.

21. If a charged particle travels through a medium of refractive index n at a speed v greater than the speed of light in the medium it emits elec-

tromagnetic waves (Cerenkov radiation) on a conical surface of half angle θ determined by

$$\cos\theta = \frac{c/n}{v}$$

See Figure 29-16 and compare with Figures 18-18 and 18-19. For an electron traveling through a medium of refractive index $n = 1.5$, calculate the emission angle θ for $v = 2.0 \times 10^8$ m/s and for $v = 2.9 \times 10^8$ m/s.

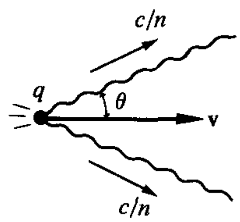


Figure 29-16

***22.** Consider, in Figure 29-17, a monochromatic electromagnetic wave characterized by the angular frequency ω and wave number k $(= \omega/c)$ traveling along the z-axis of a certain coordinate system. Let ω' and k' be the corresponding parameters for this same wave as seen by a primed observer who travels along the z-axis at the velocity u. Assuming the validity of the Lorentz transformation in (3-33) and that the phases of the wave $(kz - \omega t)$ for one observer and $(k'z' - \omega' t')$ for the other are the same, derive the Doppler shift formula

$$\nu' = \nu\left[\frac{1 - u/c}{1 + u/c}\right]^{1/2}$$

where $\nu' = \omega'/2\pi$ is the observed frequency for the primed observer. (*Hint*: Why must the relation $\omega' = k'c$ be valid for all choices for u?)

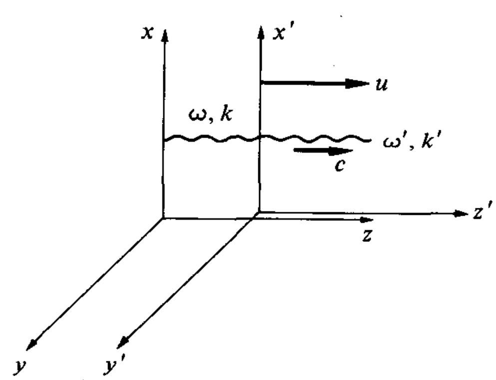


Figure 29-17

23. Using the results of Problem 22, show that if a monochromatic source emits radiation of wavelength λ, then to an observer traveling away from the source at a velocity u the radiation has the wavelength λ', given by

$$\lambda' = \lambda\left[\frac{1 + u/c}{1 - u/c}\right]^{1/2}$$

24. Light having a wavelength of 5.5×10^{-7} meter leaves a distant galaxy which is receding from us at a velocity of $c/2$. Making use of the result of Problem 23, calculate:

(a) The original frequency of this light.
(b) The wavelength of this light when it reaches earth.
(c) The observed wavelength if the galaxy were traveling toward us at a velocity of $c/2$.

25. A certain line in the spectrum of hydrogen has a wavelength, as seen on earth, of 4.35×10^{-7} meter. When this line is observed in the spectrum from a distant galaxy, it is found to have the value of 6.60×10^{-7} meter. What is the velocity of recession of this galaxy from us? What would the observed wavelength of this line be if the galaxy were approaching us at this velocity? (*Hint*: Use the Doppler shift formula in Problem 23.)

26. For a galaxy that recedes from us at a velocity of $0.99c$, what would be the observed wavelength of the 21-cm line in the spectrum of hydrogen emitted by this galaxy?

27. A spaceship that is approaching earth at a speed of $0.2c$ sends out a flash of light with a wavelength of 4.5×10^{-7} meter. What is the wavelength of this light as seen on earth?
28. Consider the solution in (29-9) for the charge-current-free Maxwellian equations.
    (a) Show that if the first relation in (29-8) is satisfied by these solutions for all closed surfaces, then so is the second.
    (b) Show that if the fourth relation in (29-8) is satisfied by these solutions for all closed curves l, then so is the third.
29. Verify that the flux out of a closed sphere of radius r and centered at the origin due to the electric field

$$\mathbf{E} = \mathbf{i}E_0 h(z - ct)$$

vanishes. (*Hint*: In terms of the angles θ and ϕ defined in Figure 29-11, the area element $d\mathbf{S}$ is given by

$$d\mathbf{S} = \mathbf{e}r^2 \sin\theta \, d\theta \, d\phi$$

where $\mathbf{e}$ is a unit vector in the radial direction.)
30. For the rectangular path in Figure 29-18 verify that the fields in (29-9) satisfy the third relation of (29-8) for all choices of the function h.

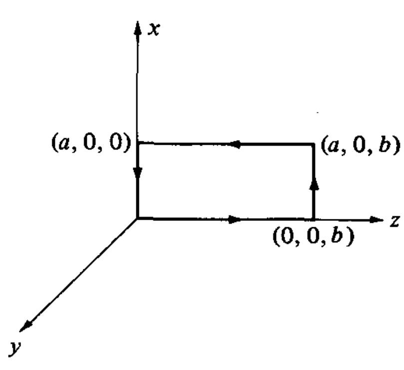


Figure 29-18

†31. A particle of charge q is moving along the trajectory

$$x = a_0 t^2 + bt$$

where a_0 and b are constants. Calculate, by use of (29-31), the rate at which it radiates energy.
†32. A particle of charge q oscillates with simple harmonic motion of angular frequency ω and amplitude A. Calculate the total energy it radiates during one complete oscillation.
†33. A particle of charge q and mass m orbits at right angles to a uniform **B**-field at a velocity v ($\ll c$).
    (a) What is its acceleration?
    (b) Show that the rate of change dW/dt of its energy is

$$\frac{dW}{dt} = -\frac{q^4 B^2}{3\pi\epsilon_0 m^3 c^3} W$$

34. Making use of the result of Problem 33, show that the fractional rate of energy loss dW/W during one period is

$$\frac{2}{3} \frac{q^3 B}{\epsilon_0 c^3 m^2}$$

and make a numerical estimate of this loss for a proton moving in **B**-field of strength 1.0 tesla.
35. For an electron traveling at a velocity of $c/137$ about a proton (a hydrogen atom) in a circular orbit of radius $a = 5.3 \times 10^{-11}$ meter calculate the total energy radiated during one complete orbit. Assume that the velocity and the orbital radius are not appreciably modified by the radiation.

30 The properties of light

We all know what light is; but it is not easy to tell what it is.

SAMUEL JOHNSON (According to Boswell)

30-1 Introduction

Visible light is the radiation associated with that part of the electromagnetic spectrum to which the human eye is receptive. Since visual perceptions play such an important role in man's daily activities, it is not surprising that the study of light, or *optics*, as it is known, began early in his history. For the earliest recorded observations of optical phenomena we are indebted to the ancient Greeks. However, it was not until well into the seventeenth century that definitive theories of the nature of light were developed.

One of the earliest quantitative attempts to describe light was made by Isaac Newton, who developed a *corpuscular theory*. Basic to this theory is the idea that the light emitted by a source consists of streams of corpuscles, which stream out in rectilinear paths in all directions from the source. One of the important implications of this theory is that the velocity of light in a dense medium, such as water, is larger than is its velocity in a less dense one such as air. As we shall see, this is *not* consistent with experiment. Mainly because of this and related incorrect predictions, the corpuscular theory of Newton was eventually abandoned. But not until the nineteenth century!

In 1670 Christian Huygens (1629–1695), who was a contemporary of Newton, proposed a competing *wave theory* of light. Although this theory was very appealing in a variety of respects, it did not receive wide acceptance for well over a century. One of the main arguments against it was that if light were indeed a wave motion, then it should exhibit wavelike phenomena, such as diffraction and interference. Since at that time diffraction effects, such as the bending of light rays around corners, had not been observed, the conclusion that light could not be a wave motion was inescapable. Today we know that the wavelenght of light is of the order of 5.0×10^{-7} meter and is thus small compared with the dimensions of macroscopic bodies. In order to observe diffraction phenomena, it is necessary to use an apparatus, such as a slit system, whose dimensions are comparable to the wavelength of visible light. Hence it was not until well over one hundred years after Huygens that the wave nature of light was unambiguously established.

Thomas Young (1773–1829) finally vindicated Huygens. He showed that when two coherent beams of light fall onto the same area of a viewing screen, there appear alternating dark and bright bands characteristic of wavelike interference phenomena. For the first time, then, man had clear evidence for the wave nature of light. Subsequently, Jean B. Foucault (1819–1868) succeeded in measuring the velocity of light in water. Contrary to the predictions of Newton, he found that the measured speed was less in water than in air. The ideas of Newton seemed to be dead. Moreover, with the wave nature of light thus established, Augustin J. Fresnel (1788–1827), assuming that light was a wave motion in the "ether," derived formulas for the intensity ratios for the reflected and the transmitted light resulting from a light beam incident on the interface between two media. These relations are known as *Fresnel's equations*. Although derived on the basis of the incorrect ether hypothesis, they have been found to be consistent with experiment nevertheless.

As we saw in Chapter 29, our views of the nature of light were modified again in the nineteenth century, when James Clerk Maxwell enunciated his equations and demonstrated that they admitted wavelike solutions corresponding to waves that travel in free space at the velocity of light c. His prediction for this speed $c = (\epsilon_0\mu_0)^{-1/2} \cong 3.0 \times 10^8$ m/s is in excellent agreement with other measurements for the speed of light. In one bold stroke, then, Maxwell showed that *all* quantitative predictions of the entire field of optics could be deduced directly from electromagnetic theory. Further, he showed that these results for visible light were applicable to a much wider range of phenomena—that is, to the entire electromagnetic spectrum, of which visible light is but a very small part.

Finally, it is interesting to note that even though the Newton–Huygens controversy regarding the nature of light seemed to have been firmly settled by Maxwell in the nineteenth century, the matter was again reopened during the early part of our own century. In an effort to account for certain

experimental results associated with *blackbody radiation* and the *photoelectric effect*, Max Planck (1858–1947) and Albert Einstein (1879–1955) hypothesized that under suitable circumstances electromagnetic waves take on the attributes of streams of corpuscles. These corpuscles are today called *photons*. Planck and Einstein suggested that a monochromatic electromagnetic wave of frequency ν and thus of wavelength $\lambda = c/\nu$ must consist of a stream of photons, each of energy E and of momentum p, with

$$E = h\nu \qquad p = \frac{h}{\lambda} \tag{30-1}$$

The constant h is known as *Planck's constant*; it has the measured value

$$h \cong 6.63 \times 10^{-34}\ \text{J-s} \tag{30-2}$$

Today we know that light and electromagnetic waves in general are capable of exhibiting both particlelike and wavelike characteristics; and that the particular aspect exhibited in a given case depends on the details of the prevailing physical situation. Certain experiments will bring out the particulate aspects of the wave, and others its wavelike features. In the following we shall be concerned only with those phenomena which exhibit the wave aspects of light. Some experiments that bring out the corpuscular nature of light—as well as the wavelike nature of elementary particles, such as electrons or protons—will be briefly discussed in Chapter 34.

We begin our study of optics in this chapter by considering first the phenomena and laws of *geometrical optics*. This study involves the reflection and refraction of light from surfaces whose dimensions are large compared to a wavelength. Under these circumstances, the fact that light is a wave motion plays only a peripheral role. Then, we shall examine certain of the phenomena of diffraction that arise when light interacts with objects of dimensions comparable to the wavelength of light. This is the field of *physical optics*. The fact that light is a wave motion plays a crucial role in the understanding of the phenomena of this field.

30-2 Measurements of the velocity of light

Prior to the time of Maxwell it was not known that light was an electromagnetic wave whose velocity c could be ascertained indirectly in terms of the electromagnetic parameters ϵ_0 and μ_0. Consequently, during this earlier period much effort was expended on direct measurements of this fundamental constant. The purpose of this section is to examine briefly some of these earlier attempts.

The first recorded effort to obtain a value for the speed of light was carried out by Galileo during the seventeenth century. Galileo and an assistant, each in possession of a lantern with a removable dark-slide, stationed themselves on two nearby hills about a kilometer apart. The experiment was then

initiated by Galileo's removing the dark slide from his lantern so that a beam of light went out towards his assistant. On seeing this light, the assistant immediately removed the slide from his own lantern, thereby sending a beam of light back toward Galileo. The time interval measured from the instant that Galileo first removed his slide to that when he saw the light from his assistant's lantern is thus a measure of the velocity of light. Unfortunately, this rather direct experiment could not succeed. Assuming that the distance d between the two lanterns is 1 km, and using the fact that the speed of light is 3.0×10^8 m/s, we find that the time interval τ for the light to go successively between the two lanterns is

$$\tau = \frac{2d}{c} = \frac{2 \times 1.0 \text{ km}}{3.0 \times 10^8 \text{ m/s}} = 6.7 \times 10^{-6} \text{ s}$$

Since this is imperceptibly small compared to a man's reaction time, a sensible value for the speed of light could not be found this way.

A second attempt to obtain a value for the speed of light achieved more success and was made in 1675 by the Danish mathematician and astronomer, Ole Roemer (1644–1710). The planet Jupiter is about five times farther away from the sun than is the earth, and orbits about the sun with a period of 11.86 years. In turn, twelve moons are in orbit about Jupiter. The innermost of these is called Io and has a period $P \cong 1.77$ days = 42.5 hours. While making measurements of this period, Roemer observed a systematic variation in its value throughout the course of the year. During that part of the year when the earth, due to its own orbital motion, recedes from Jupiter, he found the observed periods to be, on the average, larger than the values measured during the remainder of the year, when the earth approaches Jupiter. To understand how a value for c can be deduced from these data, consider this astronomical situation in Figure 30-1 (not to scale). Suppose that, at $t = 0$, Io enters Jupiter's shadow at point A. Because of the fact that the velocity of light is finite, news of this eclipse will reach earth, located at this instant at E_1, a certain time t_1 later. About 42.5 hours later, when the earth is at E_2, Io

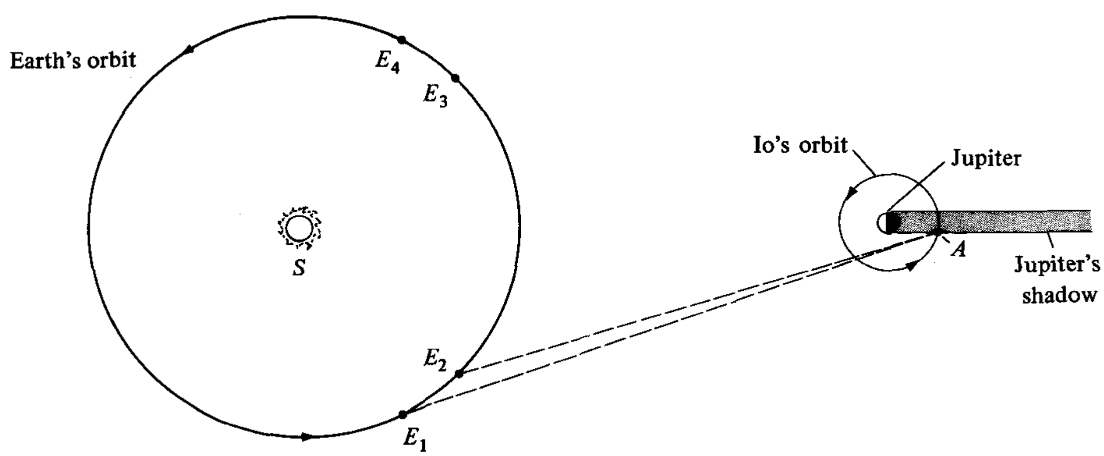


Figure 30-1

will again be eclipsed by Jupiter at A, and this event will be seen on earth at a later time, t_2. The *measured* value for the period of Io is thus $(t_2 - t_1)$. However, because of the fact that the earth travels in its orbit at a velocity of $30\ \mathrm{km/s} \cong 10^{-4}\,c$, the time it takes for the eclipse to be seen on earth the second time, when it is at E_2, is less than when it was at E_1. In other words, because of the finiteness of the speed of light, and the fact that in 42.5 hours the earth moves closer to Jupiter by the approximate distance

$$(30\ \mathrm{km/s}) \times (42.5 \times 3600\ \mathrm{s}) \cong 4.6 \times 10^6\ \mathrm{km}$$

the observed period $(t_2 - t_1)$ is less than the true value by about

$$\frac{4.6 \times 10^6\ \mathrm{km}}{3.0 \times 10^8\ \mathrm{m/s}} \cong 15\ \mathrm{s}$$

Similarly, when the earth travels from E_3 to E_4 on the other side of its orbit, the measured period will exceed the actual one by about 15 seconds.

Analyzing these data in more detail, Roemer concluded that the time required for light to travel a distance equal to the diameter of the earth's orbit is about 22 min. Since the radius of this orbit is 1.5×10^8 km, this yields for the speed of light the approximate value

$$\frac{2 \times 1.5 \times 10^8\ \mathrm{km}}{22 \times 60\ \mathrm{s}} = 2.3 \times 10^8\ \mathrm{m/s}$$

This is very close to the presently accepted value of 3.0×10^8 m/s.

A second astronomical determination of c was subsequently carried out by James Bradley (1692–1762). Bradley observed that during the course of a year a star located near the axis of the earth's orbit will appear to traverse a small circular path with a sense opposite to that of the earth's orbital motion (Figure 30-2). This phenomenon is called *aberration.*

Figure 30-3 illustrates schematically the principle underlying Bradley's measurement for the case of a star vertically above the earth's orbit. If the earth were not moving in orbit about the sun, then to a terrestrial observer the star would always be in a fixed position vertically above the orbital plane. However, the earth does travel at a certain velocity **v** about the sun. Hence, when it is at E_1, to an observer on earth the velocity of the light from

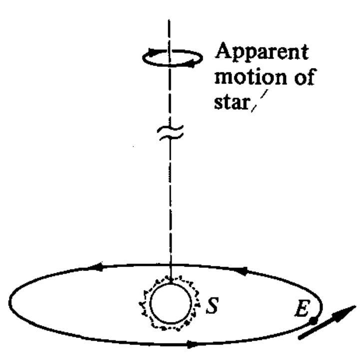


Figure 30-2

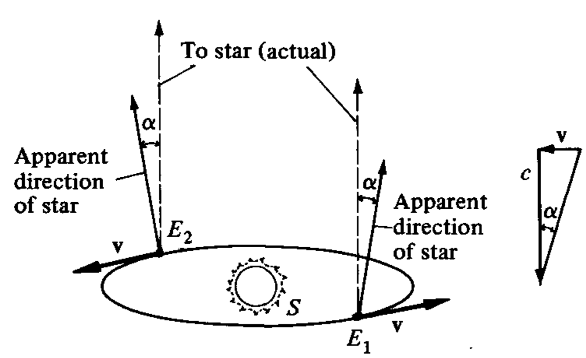


Figure 30-3

the star has a component v directed opposite to the earth's velocity. As shown in the figure, the apparent direction of the star thus makes an angle α with its actual direction, where[1]

$$\tan \alpha = \frac{v}{c}$$

Correspondingly, when the earth is on the opposite side of its orbit at E_2 the star will appear to be displaced by the angle α in the opposite direction. Hence during the course of a year the star will appear to orbit in a circle, or more accurately in an ellipse, of angular diameter 2α. Making use of the known value $v \cong 3.0 \times 10^4$ m/s and the observed value for α ($\cong 20.6''$), he obtained a value very close to the one accepted today.

An important and successful terrestrial experiment to measure c was pioneered in 1849 by Armand H. L. Fizeau (1819–1896). Figure 30-4 is a schematic diagram of the apparatus he employed. The key to this measurement is a toothed wheel, which can be set into rapid rotation about its axis. The actual measurement is initiated when the light emitted by a source S goes through a lens L_1 and is reflected toward the left by a half-silvered mirror M_1, which partially reflects and partially transmits the incident light. Assuming that at this instant the position of the wheel is such that a gap is present, the incident light will travel successively through the lenses L_2 and L_3 and reach the mirror M_2. From there it will be reflected backward and retrace its path toward the wheel. Now if the angular velocity of the wheel is such that by the time the light has traveled this distance of length $2l$ its further passage is blocked by a tooth in the wheel, then the observer O will not see any light. On the other hand, if the wheel rotates sufficiently

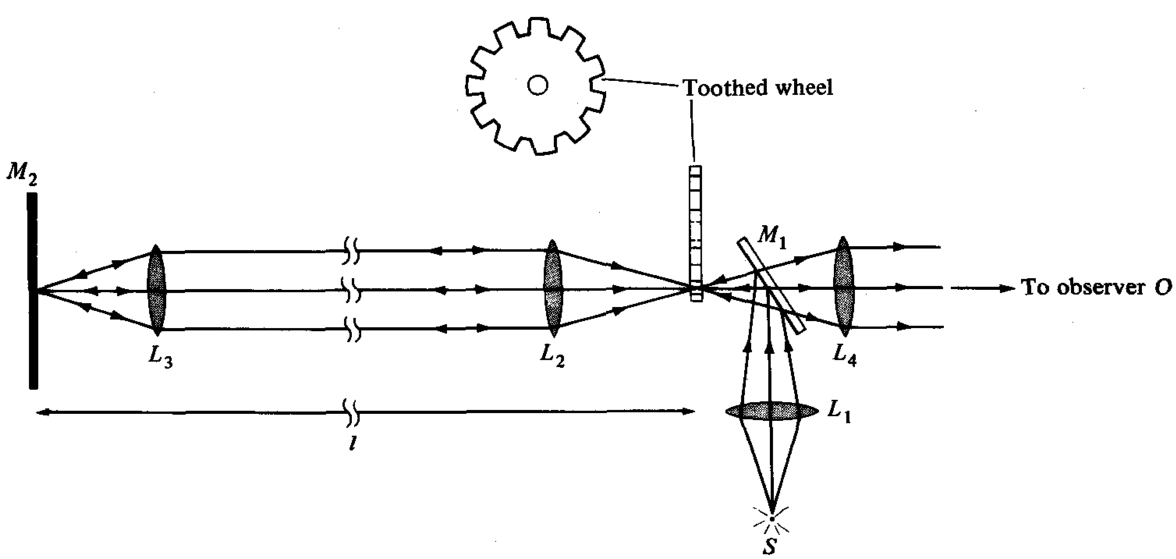


Figure 30-4

[1]According to the theory of relativity, the correct formula is: $\tan \alpha = v/(c^2 - v^2)^{1/2}$ and this reduces to Bradley's formula for $v \ll c$.

slowly, then the original light which went through a gap in the wheel will be reflected back from M_2, go through this same gap, be partially transmitted by M_1, and finally be observed at O. Thus, knowing the distance l and the angular velocity of the wheel at which no light is seen by the observer, Fizeau could measure the speed of light. Making use of a wheel with 720 teeth, and a distance $l \cong 8$ km, he found that when the wheel rotated at about 600 revolutions per minute the observer at O saw no light. He arrived at the value $c \cong 3.1 \times 10^8$ m/s.

This important experimental determination of c by Fizeau, has been followed by a number of other measurements. As a result of an ever-improving sequence of these experiments the accepted value for the velocity of light today[2] is

$$c = 2.997925 \times 10^8 \text{ m/s } (\pm 300 \text{ m/s}) \tag{30-3}$$

For most purposes the approximation $c \cong 3.0 \times 10^8$ m/s is sufficiently accurate.

30-3 Index of refraction

The value for c quoted in (30-3) refers to the speed of light only in a vacuum. The speed of light in a transparent medium, such as air, water, or glass, is, generally speaking, less than c. If v is the measured speed of light in a material medium, the *refractive index* or the *index of refraction, n,* of that medium is defined by

$$n = \frac{c}{v} \tag{30-4}$$

In the following we shall be concerned only with *nondispersive* materials, for which $n > 1$. The speed of light in such a material is always smaller than is its speed in a vacuum.

Table 30-1 lists values for the refractive index for typical substances. For

Table 30-1 Index of refraction at 293 K for $\lambda = 5.893 \times 10^{-7}$ m

Substance	*n*
Air (1 atm)	1.0003
Benzene	1.50
Crown glass	1.52
Diamond	2.42
Fused quartz	1.46
Water	1.33

[2]For a more recent and accurate value see K. M. Evenson et al., *Phys. Rev. Lett.* **29**, 1346 (1972).

most practical purposes the refractive index for air is unity, and for the other materials listed n ranges from a value of 1.33 for water to 2.42 for diamond. As we shall see below, the index of refraction varies slightly with the frequency of the light ν or equivalently with its wavelength λ given by (29-23):

$$\lambda = \frac{c}{\nu} \tag{30-5}$$

The values for n in the table correspond to the light associated with an intense line in the spectrum of sodium, which has the wavelength $\lambda = 5.893 \times 10^{-7}$ meter.

In discussions of optics one often makes use of the units of length of the micron (abbreviated μ) and the angstrom (abbreviated Å). These are defined by

$$\begin{aligned} 1 \text{ micron} &= 1\ \mu = 10^{-6}\text{ m} \\ 1 \text{ angstrom} &= 1\ \text{Å} = 10^{-10}\text{ m} \end{aligned} \tag{30-6}$$

Thus the wavelength of yellow light, $\lambda = 5893$ Å, as used in the table corresponds to a wavelength of 5.893×10^{-7} meter $= 0.5893\ \mu$. The micrometer (which is equivalent to the micron) and its abbreviation μm (see Table 1-1) are rarely used in optics.

In order to discuss the physical meaning of refractive index let us recall the relation (29-14)

$$n = (\kappa\kappa_m)^{1/2}$$

for the index of refraction n of a material characterized by a dielectric constant κ and relative permeability κ_m. As it stands, this formula is applicable throughout the electromagnetic spectrum and is not restricted to visible light. However, experiment shows that for optical frequencies ($\nu \sim 6 \times 10^{14}$ Hz), the relative permeability κ_m of most materials is, to a high degree of approximation, unity. Hence, for visible light,

$$n = \sqrt{\kappa} \tag{30-7}$$

and this suggests the following physical picture for the index of refraction. If a monochromatic electromagnetic wave of an optical frequency ν enters a dielectric medium, the electric field of the wave causes the constituent dipoles of the medium to oscillate at the same frequency ν. The resultant motion of charge produces, in turn, a secondary electromagnetic wave, also at the same frequency ν. According to the superposition property of the electromagnetic fields, the total wave traveling through the medium is the sum of the original wave and that produced by the oscillations of the dipoles. And it is the resultant *total* wave that travels through the medium at the velocity $v = c/n$, with n given in (30-7).

Based on this physical picture, the reason why the index of refraction might vary with frequency is readily apparent. For the efficiency with which the dipoles of a given medium oscillate generally varies with the frequency of the incident wave. It follows from (30-7) that the index of refraction will

also vary with the wavelength or the frequency of the incident wave. Figure 30-5 shows a plot of the observed values of the refractive index of crown glass as a function of wavelength at a fixed temperature. Note that this variation is *not* very large. Indeed, experiment shows that for many substances this variation can be safely neglected and we shall assume this to be so in much of the following.

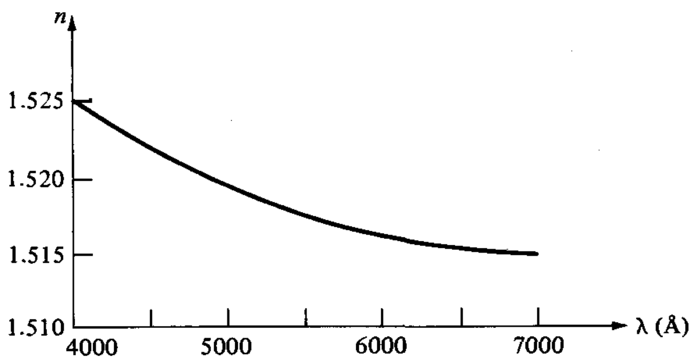


Figure 30-5

In a similar way, the above physical picture for refractive index also suggests and makes plausible the fact that n might vary with temperature. By way of example, in Table 30-2 we list the values of n for water at a wavelength of 0.5893 μ at several temperatures. Again, experiment shows that for many substances the temperature variation of n can (within reasonable limits) be safely neglected and we shall assume this to be so in the following.

Table 30-2 Index of refraction of water for λ = 5893 Å

T (°C)	15	25	35	45	55
n	1.3338	1.3329	1.3316	1.3301	1.3285

Example 30-1 A monochromatic wave of yellow light of wavelength 5893 Å is traveling in free space.

(a) What is the frequency of this wave?
(b) What is its velocity in a medium for which n = 1.52?
(c) What is its wavelength in the medium?

Solution

(a) Substituting the given value for λ into (30-5) we find that

$$\nu = \frac{c}{\lambda} = \frac{3.0 \times 10^8 \text{ m/s}}{5893 \text{ Å}} = \frac{3.0 \times 10^8 \text{ m/s}}{5.893 \times 10^{-7} \text{ m}}$$
$$= 5.1 \times 10^{14} \text{ Hz}$$

where in the last equality only two significant figures have been kept.

(b) Making use of the definition of n in (30-4) we find, for the wave speed v in the medium,

$$v = \frac{c}{n} = \frac{3.0 \times 10^8\ \text{m/s}}{1.52} = 2.0 \times 10^8\ \text{m/s}$$

(c) Since the frequency of the wave is the same in free space as in the medium, it follows that the wavelength λ_m in the medium is less than its value λ_0 in free space by the factor n^{-1}. See (29-24). Hence

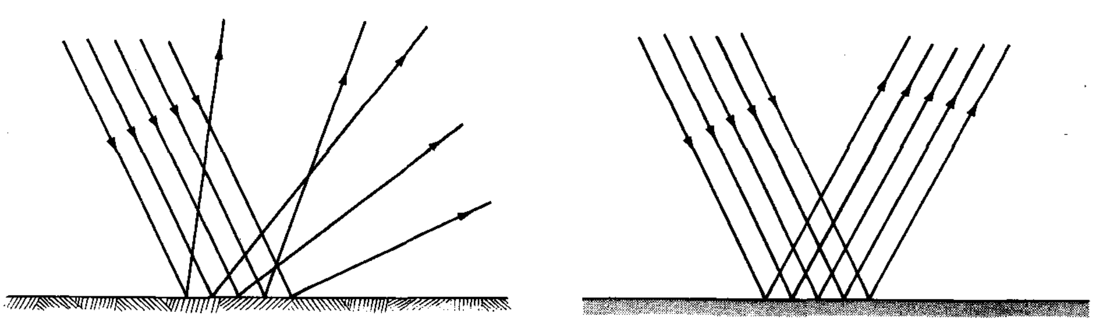
$$\lambda_m = \frac{\lambda_0}{n} = \frac{5893\ \text{Å}}{1.52} = 3900\ \text{Å}$$

to two significant figures.

30-4 The laws of geometrical optics

If a beam of visible light falls on the surface of an object such as a piece of wood, the reflected light goes out in all directions, and the object can be seen from several directions. This type of reflection is called *diffuse reflection* and is exemplified in Figure 30-6a. Diffuse reflection arises if the normal irregularities associated with the surfaces of most objects are very large compared to the wavelength of the light being used. If a beam of light falls on a very smooth or highly polished surface—that is, one for which the irregularities are small compared to the wavelength of light—then the reflected light goes out as a beam in a single direction. This type of reflection is called *specular* and is shown in Figure 30-6b. For example, the reflection of light from the wall or floor of a room is generally diffuse, whereas that reflected from a mirror is specular. In the remainder of this chapter we shall be concerned only with specular reflection. Hence it will be assumed that the irregularities of all surfaces of interest are small compared to the wavelength of visible light, and thus in the following the word "reflection" shall always mean "specular reflection."

Consider, in Figure 30-7, a beam of monochromatic light, which approaches, at an angle θ_1 with respect to the normal, the planar interface



(a) (b)

Figure 30-6

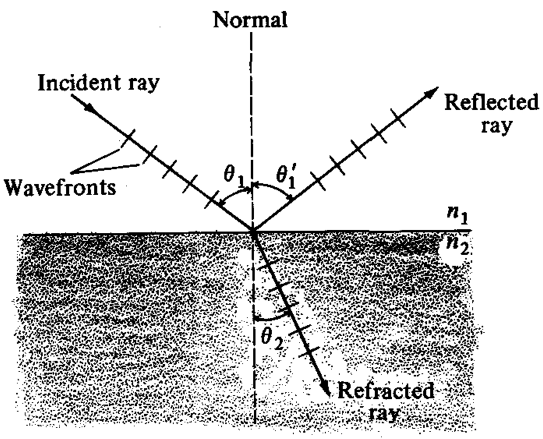


Figure 30-7

between two media of respective indices of refraction n_1 and n_2. The incoming beam is called the *incident ray*, and planes at right angles to the direction of propagation of this ray are called the *wavefronts* associated with the wave. Experiment shows that, as a consequence of the incident ray's striking the interface, there arise two additional waves. One of these propagates backward into the same medium as the incident wave and is called the *reflected wave.* It propagates along the direction of the *reflected ray* at some angle θ_1' with respect to the normal. The second wave arising at the interface propagates into the lower medium and it is known as the *refracted wave.* It propagates along the direction of the *refracted ray*, which makes a certain angle θ_2 with respect to the normal. Associated with the reflected and the refracted waves there are also wavefronts, and these are represented in the figure by short lines perpendicular to the respective rays. Since the electric and magnetic vectors of an electromagnetic wave are perpendicular to the direction of propagation of the wave, it follows that the wavefronts of the various waves are the planes containing the electric and the magnetic vectors.

Experiment as well as detailed studies of Maxwell's equations show that the intensities and the directions of propagation of the incident, the reflected, and the refracted rays in Figure 30-7 are related to each other in a certain way. The laws of *geometrical optics* relate the directions of propagation of these three rays and are:

1. The incident, the reflected, and the refracted rays all lie in a single plane. This plane is called the *plane of incidence* and can be thought of as being determined by the incident ray and the normal to the interface.
2. The *angle of reflection* θ_1' is equal to the *angle of incidence* θ_1:

$$\theta_1 = \theta_1' \tag{30-8}$$

3. The angle of incidence θ_1 and the *angle of refraction* θ_2 are related by *Snell's law*:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{30-9}$$

Although we shall be concerned with these three laws of geometrical optics only for visible light, it should be noted that they are applicable very generally for all electromagnetic waves.

It is interesting that although these laws can be derived from Maxwell's theory, they were known much earlier. Thus the law of reflection in (30-8) was known to Euclid and the law of refraction was determined empirically in the seventeenth century by Willebroad Snell (1591–1626). Snell's law was also derived from Newton's corpuscular theory by René Descartes (1596–1650). However, the fact that these laws are applicable to the entire electromagnetic spectrum was not known until late in the nineteenth century.

30-5 Applications

In this section we apply the laws of geometrical optics to a number of particular physical situations.

Example 30-2 A directed beam of light from a source S strikes a mirror (that is, an interface for which most of the incident light is reflected) at a point A and at an angle of 60° with respect to the normal N_1; see Figure 30-8. Assuming that there is a second mirror inclined at an angle of 130° with respect to the first, find the three angles θ_1, θ_2, and θ_3 as defined in the figure. Assume that N_2 is the normal to the second mirror.

Solution According to the law of reflection, the angle of incidence is equal to the angle of reflection. Hence

$$\theta_1 = 60°$$

To obtain θ_2 and θ_3, note that in triangle ABC the measure of angle B is 130° while that for angle CAB is $90° - \theta_1 = 30°$. Hence, since the sum of the angles of a triangle is 180°, it follows that

$$\theta_2 = 180° - 130° - 30° = 20°$$

Finally, since N_2 is the normal to the second mirror, a second application of the law of reflection yields

$$\theta_3 = 70°$$

Example 30-3 A ray of light traveling in air strikes, at an angle of incidence of 45°, the planar surface of a piece of glass of index $n = 1.5$. Calculate the angle of refraction θ (Figure 30-9).

Solution Since the refractive index of air is unity, Snell's law in (30-9) takes the form

$$\sin 45° = 1.5 \sin \theta$$

where we have made the substitutions $n_1 = 1$, $n_2 = 1.5$, $\theta_1 = 45°$, and $\theta_2 = \theta$. Hence

$$\sin \theta = \frac{\sin 45°}{1.5} = 0.47$$

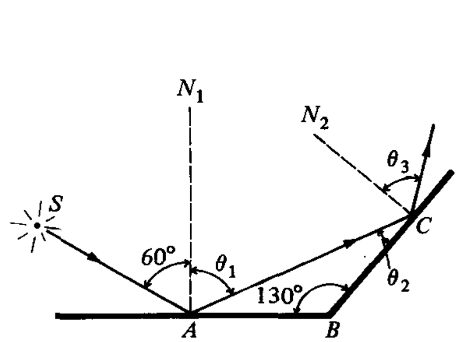


Figure 30-8

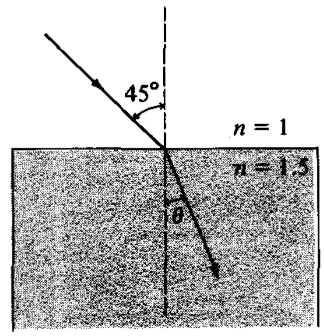


Figure 30-9

which, by reference to a table of trigonometric functions, is equivalent to

$$\theta \cong 28°$$

Example 30-4 A beam of light is produced by a source S, which is under water and at a vertical distance d below the surface, as in Figure 30-10 (not to scale). Assuming almost normal incidence so that the angles of incidence and refraction θ_1 and θ_2, respectively, are sufficiently small so that $\sin\theta_1 \cong \tan\theta_1 \cong \theta_1$, and similarly for θ_2, find the depth y below the surface where the source appears to be to the observer O. Assume that $n = 4/3$.

Solution Since the incident ray is under water and the refracted ray is in air, according to Snell's law the angles θ_1 and θ_2 are related by

$$\frac{4}{3}\sin\theta_1 = \sin\theta_2 \tag{30-10}$$

By hypothesis, the angles θ_1 and θ_2 are presumed small, and thus

$$\sin\theta_1 \cong \tan\theta_1 = \frac{a}{d}$$

$$\sin\theta_2 \cong \tan\theta_2 = \frac{a}{y}$$

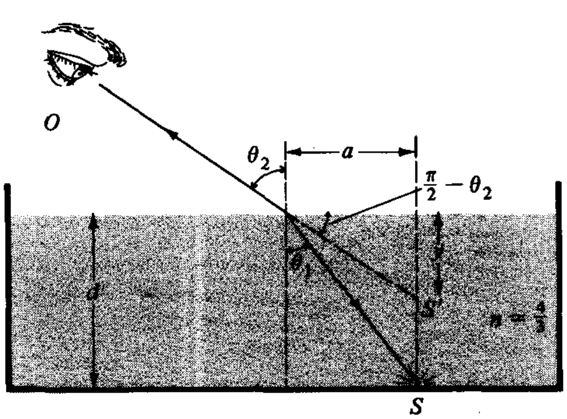


Figure 30-10

Substitution into (30-10) then yields

$$\frac{4a}{3d} = \frac{a}{y}$$

or, equivalently,

$$y = \frac{3}{4}d$$

In other words, to the observer O, the depth of the water appears to be only three fourths of its actual depth.

More generally, if a flat pan or a lake or a stream is filled to a depth d with a transparent substance of refractive index n, then its apparent depth to an observer looking perpendicularly down on the surface is d/n.

Example 30-5 Consider, in Figure 30-11, a ray incident, at an angle θ_1, on a face of a prism of vertex angle A and of index of refraction n. Find the angle of deviation δ of this ray as it traverses the prism in terms of the parameters n, θ_1, and A.

Solution From the fact that the exterior angle of a triangle is equal to the sum of the remote interior angles it follows that

$$\delta = \theta_1 - \theta_1' + \theta_2 - \theta_2'$$

Further, since the sum of the angles θ_1' and θ_2' is supplementary to angle B and the latter is the supplement of angle A we have the additional relation

$$A = \theta_1' + \theta_2'$$

Hence applying Snell's law at the two refractive surfaces

$$\sin\theta_1 = n\sin\theta_1'$$
$$\sin\theta_2 = n\sin\theta_2'$$

and eliminating the angles θ_1', θ_2, and θ_2' from among the above four relations we obtain finally

$$\delta = \theta_1 - A + \sin^{-1}[(n^2 - \sin^2\theta_1)^{1/2}\sin A - \cos A\sin\theta_1]$$

An analysis of this formula shows that for any given prism there is a value for the angle of incidence θ_1 for which this deviation angle δ is a minimum. It is left as an

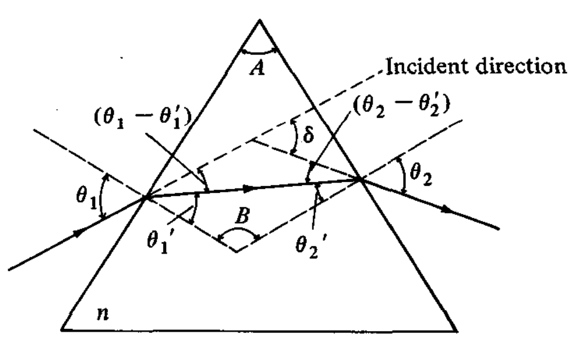


Figure 30-11

exercise to show that this minimum value for δ, call it δ_m, satisfies

$$n \sin\frac{A}{2} = \sin\frac{A+\delta_m}{2} \tag{30-11}$$

and that the value of the angle of incidence, call it $\bar{\theta}_1$, at which this occurs is

$$\bar{\theta}_1 = \sin^{-1}\left[n \sin\frac{A}{2}\right]$$

This formula for the angle of minimum deviation δ_m in (30-11) is often used when it is desired to measure the refractive index of transparent solids. The minimum angle of deviation δ_m for any given prism can be measured by varying the angle of incidence θ_1 and observing the varying direction of the transmitted ray. Combining the value of δ_m so obtained with a knowledge of the angle A we can obtain by use of (30-11), a measured value for n. By varying the color or the wavelength of the incident light, a dispersion curve such as that in Figure 30-5 can be constructed.

30-6 Total internal reflection

Suppose, in Figure 30-7, that the incident ray is in air, and therefore $n_1 = 1$. Then, the angle θ_2 of the refracted ray will be less than the angle of incidence θ_1, or, in other words, the refracted ray is bent toward the normal. For according to Snell's law

$$\sin\theta_1 = n_2 \sin\theta_2 \tag{30-12}$$

and, since $\sin\theta$ is an increasing function of θ for $0 \leq \theta \leq 90°$ we find, assuming $n_2 > 1$ that, except at normal incidence,

$$\theta_2 < \theta_1$$

Imagine now an experiment in which the angle of incidence θ_1 is increased to its maximum value of 90°. The angle of refraction will also increase and will approach a certain maximum value, call it θ_m. Substituting the value $\theta_1 = 90°$ into (30-12) we find that the maximum angle of refraction θ_m satisfies

$$\sin\theta_m = \frac{1}{n_2} \tag{30-13}$$

In other words, regardless of the angle of incidence, all refracted light rays coming from above the surface in Figure 30-7 are confined to the interior of a cone of half-angle θ_m as defined in (30-13). For example, since the index of refraction of water is 4/3, it follows that if you are under water and look upward, all of the light coming to you from above is confined to the interior of a cone of half-angle $\theta_m = \sin^{-1} 0.75 = 49°$.

Consider now the situation depicted in Figure 30-12 in which a ray of light originally travels in a medium of refractive index $n_1 > 1$ and then enters a second medium, such as air, which has an index of unity. If θ_1 is, as before, the angle of incidence and θ_2 the corresponding angle of refraction, then

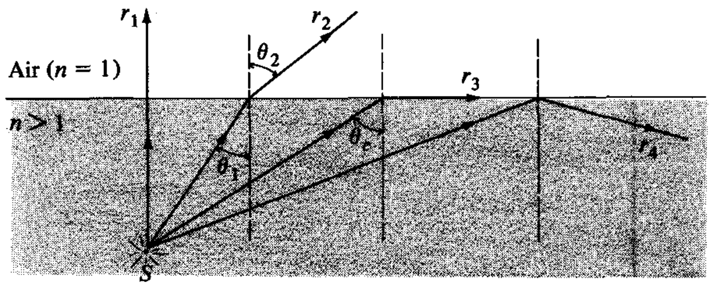


Figure 30-12

according to Snell's law

$$n_1 \sin \theta_1 = \sin \theta_2 \tag{30-14}$$

This time, as exemplified by the ray r_2 in Figure 30-12, the angles θ_1 and θ_2 are related by

$$\theta_2 > \theta_1$$

In other words, for this case, the angle of refraction (except for the normal ray r_1) is invariably larger than is the angle of incidence and thus the refracted rays are all bent away from the normal.

Imagine now an experiment involving light going from such a refractive medium into air. As the angle of incidence θ_1 is increased, it will eventually achieve a certain value, θ_c, known as the *critical angle* for which the angle of refraction $\theta_2 = 90°$. This is represented by the ray r_3 in the figure. Substituting this value $\theta_2 = 90°$ into (30-14) we find for θ_c

$$\sin \theta_c = \frac{1}{n_1} \tag{30-15}$$

The physical significance of θ_c is that for all angles of incidence θ_1 in the range $\theta_c < \theta_1 < 90°$ there exists no real angle of refraction θ_2 that satisfies Snell's law. Hence there can be no refracted ray for such angles. According to the discussion in Chapter 29, however, there is a flow of energy associated with the incident, the reflected and the refracted ray. To be consistent with the law of energy conservation, then, it follows that for $\theta_1 > \theta_c$, for which there is no refracted ray, *all* of the energy incident on the interface must be reflected back into the medium. This phenomenon is therefore called *total internal reflection.* Although there is generally some reflection at all angles, only for $\theta_1 > \theta_c$ is the reflection total.

It should be emphasized that total internal reflection is associated only with a beam of light going from a medium of higher to one of lower index of refraction. Generally speaking, a refracted ray is always associated with a light ray that is incident from the lower index side onto the interface between two media.

Example 30-6 Calculate the critical angle θ_c associated with an interface between air and:

(a) Water, of refractive index 1.33.
(b) Glass, of refractive index 1.5.
(c) Diamond, of refractive index 2.5.

Solution In each case, making use of (30-15), we find that

(a) For water,

$$\theta_c = \sin^{-1} 0.75 = 49°$$

(b) For glass,

$$\theta_c = \sin^{-1} 0.67 = 42°$$

(c) For diamond,

$$\theta_c = \sin^{-1} 0.4 = 24°$$

Example 30-7 A beam of light is incident from the glass side of the interface between water ($n = 4/3$) and glass ($n = 3/2$), as in Figure 30-13. For what angles of incidence will this beam experience total internal reflection?

Solution If θ_1 is the angle of incidence and θ_2 the angle of refraction, then, according to Snell's law,

$$\frac{3}{2} \sin \theta_1 = \frac{4}{3} \sin \theta_2$$

The critical angle θ_c is defined, as above, to be the value of θ_1 for which $\theta_2 = 90°$. Hence

$$\frac{3}{2} \sin \theta_c = \frac{4}{3}$$

or, equivalently,

$$\theta_c = \sin^{-1} \frac{8}{9} = 63°$$

Thus, total internal reflection takes place for all incident angles θ in the range

$$63° \leqslant \theta \leqslant 90°$$

Example 30-8 A beam of light is incident on a 45°-45°-90° prism, of index 1.5; see Figure 30-14. Describe what happens to the beam if it is incident normally on:

(a) A leg of the prism as in Figure 30-14a.
(b) The hypotenuse of the prism as in Figure 30-14b.

Solution Since the refractive index is 1.5 it follows that with respect to air the critical angle θ_c is

$$\theta_c = \sin^{-1} \frac{1}{n} = \sin^{-1} 0.67 = 42°$$

(a) If the beam is incident normally on a leg of the prism, then the angle of incidence at point A is 45°. Since this exceeds the critical angle, it follows that the incident beam will be totally internally reflected. Accordingly, the ray follows the

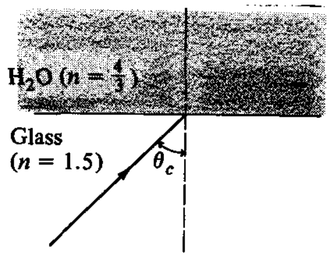


Figure 30-13

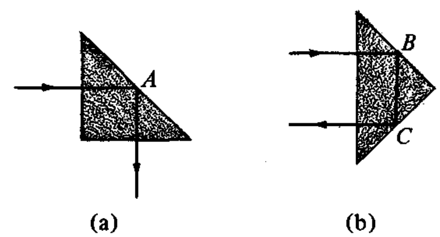


Figure 30-14

path indicated in Figure 30-14a. In effect, the direction of the beam is turned through 90°.

(b) In a similar way, if the incident beam strikes the hypotenuse of the prism as in Figure 30-14b, then at both B and C its angle of incidence is 45°. This exceeds the critical angle $\theta_c = 42°$, so the beam undergoes total internal reflection at both of these points. In effect, the direction of the beam is reversed.

30-7 Fresnel's equations

Consider again Figure 30-7, in which a light beam strikes the interface between the media of refractive indices n_1 and n_2. Maxwell's equations enable us to predict not only the directions of the reflected and the refracted beams, but also their intensities relative to that of the incident beam. The amplitudes of these relative intensities satisfy certain relations known as *Fresnel's equations.*

As we saw in Chapter 29, a monochromatic electromagnetic wave traveling in free space, is characterized uniquely by its amplitude E_0 and wavelength λ. (See (29-20) through (29-22).) The average flow of energy $\bar{N}$ per unit area per unit time associated with such a wave when traveling in free space is given by

$$\bar{N} = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\,E_0^2 \qquad \textbf{(29-25)}$$

and thus depends exclusively on the amplitude E_0. More generally, if an electromagnetic wave travels in a (nonmagnetic) medium of index of refraction n, then the argument used in Example 29-4 shows that the relation (29-25) is still applicable, provided that the factor ϵ_0 is replaced by $\epsilon_0\kappa$. Hence, introducing the symbol I to represent the energy flow per unit area per unit time associated with a monochromatic wave traveling through a medium of refractive index n, we find by use of (29-14) that

$$I = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\,nE_0^2 \qquad \textbf{(30-16)}$$

with E_0 the amplitude of the electric field vector of the wave. In the following, we shall use the term *intensity* for the quantity I given by (30-16).

Consider, in Figure 30-15, an electromagnetic wave of amplitude E_0

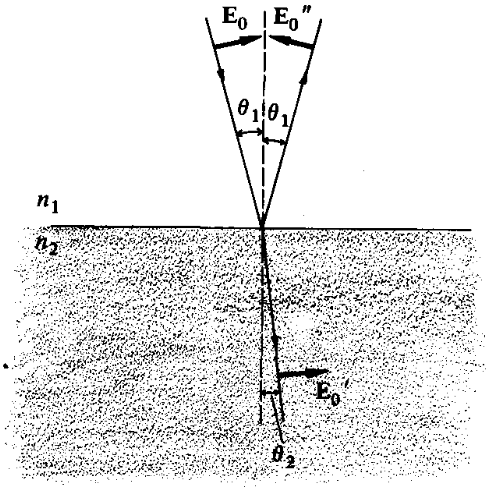


Figure 30-15

traveling in a medium of refractive index n_1 and striking at an angle θ_1 the interface with a second medium of refractive index n_2. Let E_0'' be the corresponding amplitude of the reflected wave and E_0' that of the refracted wave. The directions of the electric vectors for these three waves must be perpendicular to the respective propagation directions, and for the directions of the electric fields shown in the figure the associated **B** vectors are perpendicular to and out of the plane of the diagram. According to (30-16), the energy flow associated with the incident, the reflected, and the refracted rays are completely specified in terms of the respective amplitudes E_0, E_0', and E_0''.

The relations between these three amplitudes E_0, E_0', and E_0'' are in general fairly complex and thus we shall consider in detail only the special case of near-normal incidence. Here the angle of incidence satisfies the relation $\theta_1 \ll 1$ and thus Snell's law has the form

$$n_2\theta_2 = n_1\theta_1$$

In the problems it will be verified that for these angles, the ratios E_0''/E_0 of the reflected to the incident amplitude and E_0'/E_0 of the refracted to the incident amplitudes are given by

$$\frac{E_0''}{E_0} = \frac{n_2 - n_1}{n_1 + n_2} \qquad \frac{E_0'}{E_0} = \frac{2n_1}{n_1 + n_2} \tag{30-17}$$

These are *Fresnel's equations* for this special case of normal incidence. Substitution into (30-16) yields, for the reflected intensity I'' and the refracted intensity I',

$$\begin{aligned} I'' &= I_0\left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 \\ I' &= I_0\frac{4n_1n_2}{(n_1 + n_2)^2} \end{aligned} \tag{30-18}$$

where I_0 is the intensity of the incident light and where, for example, the second equality follows since, according to (30-16), $I'/I_0 = n_2E_0'^2/n_1E_0^2$.

For the special case in which $n_2 = n_1$, we expect no reflected light since, in effect, there is no interface. Substitution of the value $n_2 = n_1$ shows that, consistent with this expectation, the relations in (30-18) imply that all of the incident light is transmitted: ($I'' = 0$; $I' = I_0$). If the two relations in (30-18) are added together, we find after some algebra that

$$I'' + I' = I_0$$

so that regardless of the values for n_2 and n_1 the sum of the reflected and refracted intensities is equal to the incident intensity. In this sense the intensity ratios in (30-18) are consistent with the ideas of energy conservation, as they must be.

More generally, if the angle of incidence θ_1 is not small, then if the light is polarized with the electric vectors in the plane of incidence, as in Figure 30-15 (p-polarization), then (30-17) must be replaced by

$$\frac{E_0''}{E_0} = \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \qquad \frac{E_0'}{E_0} = \frac{2\cos\theta_1 \sin\theta_2}{\sin(\theta_1 + \theta_2)\cos(\theta_1 - \theta_2)} \tag{30-19}$$

with θ_2 given by Snell's law. Correspondingly, if the light is polarized with the electric vectors perpendicular to the plane of incidence (s-polarization), then the ratios of these amplitudes are

$$\frac{E_0''}{E_0} = \frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \qquad \frac{E_0'}{E_0} = \frac{2\cos\theta_1 \sin\theta_2}{\sin(\theta_1 + \theta_2)} \tag{30-20}$$

These relations in (30-19) and (30-20) are known as Fresnel's equations.

It is established in the problems that, for normal incidence ($\theta_1 \approx 0$ and $\theta_2 = n_1\theta_1/n_2$), both (30-19) and (30-20) reduce to (30-17). This must follow since the two cases of the incident ray being polarized in, or perpendicular to, the plane of incidence are physically indistinguishable in this case.

Example 30-9 A beam of light traveling in air is incident on a glass plate of refractive index 1.5. If *all* of the light is transmitted, and the electric vector lies in the plane of incidence, as in Figure 30-15, what is the angle of incidence θ_1?

Solution Reference to (30-19) shows that since $\tan 90° = \infty$ the reflected amplitude E_0'' will vanish if $\theta_1 + \theta_2 = 90°$. Substituting this datum and the values $n_1 = 1$ and $n_2 = 1.5$ into Snell's law, we find that

$$\sin\theta_1 = n_2 \sin\theta_2 = 1.5 \sin(90° - \theta_1) = 1.5 \cos\theta_1$$

Hence, $\tan\theta_1 = 1.5$, and this yields

$$\theta_1 = 56°$$

As will be discussed in Chapter 33, this angle of incidence at which no reflection takes place is known as *Brewster's angle.*

30-8 Summary of important formulas

If a light beam traveling in a medium of refractive index n_1 is incident at an angle θ_1 on the interface between it and a second medium of refractive index n_2, then some of the light will be reflected at the same angle θ_1 and the rest will be refracted into the second medium at an angle θ_2. The angle of refraction θ_2 is given by Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{30-9}$$

Near normal incidence ($\theta_1 \cong \theta_2 n_2/n_1$) the intensities associated with the incident beam, I_0, the transmitted beam I', and the reflected beam I'' are related by

$$\begin{aligned} I'' &= I_0\left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 \\ I' &= I_0\frac{4n_1n_2}{(n_2 + n_1)^2} \end{aligned} \tag{30-18}$$

QUESTIONS

1. Define or describe briefly what is meant by the following terms: (a) aberration of starlight; (b) index of refraction; (c) critical angle; (d) Fresnel's equations; and (e) specular reflection.
2. State an experimental fact that is in contradiction with Newton's corpuscular theory. Are the laws of reflection and refraction consistent with this theory?
3. In what way or ways are the photons of Einstein and Planck different from Newton's corpuscles? In what way or ways are they similar?
4. Define what is meant by the "wavelength" of light and make use of your definition to distinguish between the phenomena of geometrical and of physical optics.
5. What is the physical basis for the difference between diffuse and specular reflection?
6. Light from a flashlight is reflected from a mirror in a dark room. Explain why, even though the reflection is specular, the image of the light can be seen from various parts of the room.
7. How far apart would Galileo and his assistant have had to be in order that the time for the light to travel back and forth between them be 1 second? How does this compare to the distance between the earth and the moon?
8. Figure 30-16 shows a curved, transparent rod made of a material of high

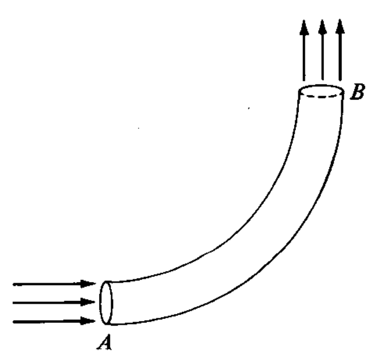


Figure 30-16

refractive index. By use of the notion of total internal reflection and Fresnel's equations, explain why it is that if light is admitted at one end, A, of the rod, an appreciable fraction of this light might emerge at the other end, B. (*Note*: When used in this way, the rod is often referred to as a *light pipe*, and the field of study concerned with the optical properties of a collection of flexible light pipes is known as *fiber optics*.)

9. Is the concept of refractive index restricted to visible light or is it more generally applicable? Explain.
10. Can Snell's law be applied to X rays? To ultraviolet light? Explain.
11. The wavelength of green light is 5600 Å in air and its wavelength in water is 4200 Å. If you view this light while your head is under water, would it appear to be green or violet? Explain.
12. On what basic physical laws are the laws of reflection and refraction based? How were they originally discovered?
13. By reference to the graph in Figure 30-5, determine whether or not the speed of light of wavelength 6000 Å in crown glass is greater or less than the speed of light of wavelength 4000 Å in the same medium.
14. Suppose that white light—that is, light containing a mixture of all visible wavelengths—is incident on the prism in Figure 30-11. Assuming that the index of a refraction is a decreasing function of wavelength, the various colors in the incident beam will come out at various angles θ_2. Will violet light correspond to a smaller or larger angle of deviation δ?
15. You are given a prism of a certain index n and of vertex angle 30°, as in Figure 30-11. Describe how you would measure the variation of n with wavelength by its use.
16. Light is incident from air onto the face of a medium of refractive index 2. What is the maximum possible angle of refraction?
17. What is the critical angle associated with the system in Question 16? Is there a relation between the critical angle and the angle computed in Question 16? Explain.
18. Describe what you would see if you were under water and looked vertically upward. Assume that $n = 4/3$ and that the surface of the water is smooth.
19. Explain why a swimming pool that is actually 2.5 meters deep appears to be only about 2.0 meters deep to an observer looking perpendicularly down on the surface.
20. In terms of the properties of the microscopic constituents of a dielectric, what is the physical mechanism that underlies the phenomena of reflection and refraction? Would reflection and refraction take place for a light beam incident on a transparent magnetic material?
21. Consider a monochromatic light beam incident normally on the interface between two media. Assuming that $n_1 \neq n_2$, is it possible for all of the incident light to be transmitted? For it all to be reflected?
22. A light beam consisting of more than one color is incident from air onto a plate of crown glass, whose refractive index is plotted in Figure 30-5. Explain why there is a separation of the colors in the refracted beam but not in the reflected beam.

PROBLEMS

1. (a) Calculate the time it takes for light to reach us from the sun. Assume the earth's orbital radius to be 1.5×10^8 km.
   (b) What are the maximum and minimum times it takes light to reach us from the planet Jupiter? Assume that Jupiter's orbit has a radius of 7.8×10^8 km.
2. Consider a beam of orange light of wavelength 6500 Å.
   (a) What is its frequency?
   (b) What is its wavelength when traveling in crown glass of index 1.52?
   (c) What is its velocity in this medium?
3. If the speed of light in a certain medium is 2.0×10^8 m/s what is the refractive index of the medium? What is the wavelength of this light in the medium if its free-space wavelength is 4500 Å?
4. Light of wavelength 5000 Å enters a medium where its wavelength is found to be 3500 Å.
   (a) What is the refractive index of the medium?
   (b) What is the frequency of the light?
   (c) What is the dielectric constant of the medium at this frequency?
5. A beam of light is incident at an angle θ onto the surface of a plane mirror. If the mirror is rotated by an angle α about an axis in its plane and perpendicular to the plane of incidence, by what angle is the reflected beam rotated?
6. By use of the graph in Figure 30-5 calculate the ratio of the speed of light in crown glass for light of wavelength 7.0×10^{-7} meter to light of wavelength 4.0×10^{-7} meter.
7. In Figure 30-17 suppose that light goes from point A to point B after reflection by mirror M. Show that if the path of the ray is such that the angle of incidence is equal to the angle of reflection, then the time required for the light to travel from A to B is a minimum. (*Hint*: Show that the time $t(x)$ to go via the indicated path is

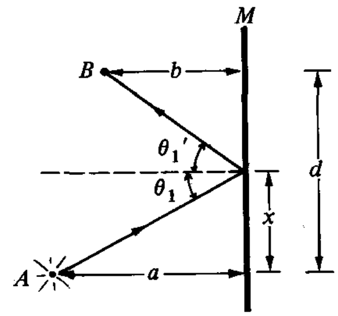


Figure 30-17

$$t(x) = \frac{1}{c}\{[a^2 + x^2]^{1/2} + [(d-x)^2 + b^2]^{1/2}\}$$

and find the value of x for which this is minimum.) (*Note*: This and the result of Problem 8 represent a derivation of the laws of geometrical optics by use of *Fermat's principle*, which states that the actual path—out of the totality of possible paths—followed by a light beam is that one which takes the *least time*.)
8. Consider, in Figure 30-18, a beam of light that goes from point A to point B inside a medium of refractive index n. If the parameters a, b, and d are as defined in the figure:
   (a) Show that the time t required to go from A to B is

$$t = \frac{1}{c}\{[a^2 + x^2]^{1/2} + n[b^2 + (d-x)^2]^{1/2}\}$$

   (b) Find that value of x for which the time t is a minimum and

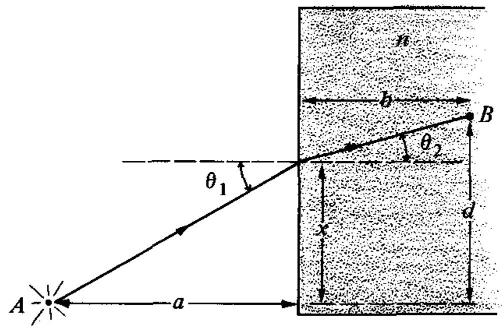


Figure 30-18

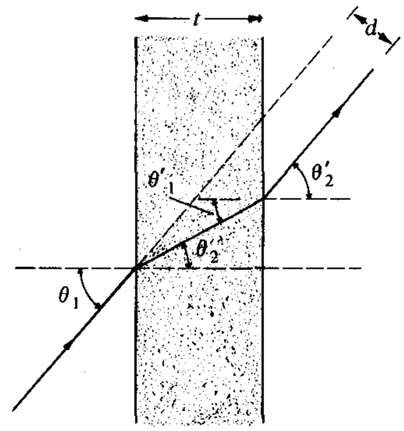


Figure 30-19

show that at this condition the angles θ_1 and θ_2 are related by Snell's law.

9. A beam of light originates at a point below the surface of a tank of water. Assuming that $n = 1.33$, calculate the angle the emerging beam makes with respect to the vertical if the angle of incidence is: **(a)** 30°; **(b)** 45°; **(c)** 75°.
10. For the physical situation described in Problem 9 calculate the angle of incidence for which the emerging beam will come out horizontally. What happens if the angle of incidence is increased beyond this value?
11. A plate of glass (index 1.5) is under water (index 1.33).
    (a) If light is originally traveling in the water and is incident at an angle of 30°, what is the angle of refraction?
    (b) If a light beam is originally traveling in the glass and is incident at an angle of 60°, what is the angle of refraction as the beam enters the water?
12. A beam of light is incident at an angle θ_1 onto a flat glass slab of thickness t and refractive index n; see Figure 30-19.
    (a) What is the value of θ_2 in terms of n and θ_1?
    (b) Why must $\theta_1' = \theta_2$ and $\theta_2' = \theta_1$?
    (c) Show that the lateral deviation d of the beam after the second refraction is

$$d = t \sin\theta_1\left[1 - \frac{\cos\theta_1}{(n^2 - \sin^2\theta_1)^{1/2}}\right]$$

and evaluate this for small θ_1.

13. A ray of light is incident at an angle of 30° on a glass plate 1 cm thick. If the lateral displacement between the incident and transmitted beams is 0.2 cm, calculate the index of refraction of the plate by use of the results of Problem 12.
14. Show that if a beam of light is incident from air on a surface of refractive index n at an angle θ_0, so that the reflected and refracted beams are perpendicular to each other, then

$$\tan\theta_0 = n$$

15. Two parallel beams of light go through the prism in Figure 30-11 of vertex angle $A = 30°$. If the refractive index of the prism for one of the beams is 1.50 and for the other it is 1.60, and if the beams enter at an angle of incidence corresponding to the minimum deviation for the former, calculate:
    (a) The deviation δ of each beam.
    (b) The angle between the two beams when they emerge from the prism.
16. Show that the angle of minimum

deviation δ_m for a prism of vertex angle A is given by (30-11).

17. Consider again, in Figure 30-14a, a beam of light incident normally onto one of the shorter sides of a 45°-45°-90° prism. What is the minimum value for the refractive index so that the beam undergoes total internal reflection? Assume the prism to be in air.

18. Repeat Problem 17, but assume this time that on the surface of the prism at A there is a thin layer of water of refractive index 1.33.

19. Light is incident at an angle θ_1 on the face of a cube of glass of index 1.5; see Figure 30-20. Find the maximum value for the angle θ_1 so that the light is totally internally reflected at point A on an adjacent face. Assume that the cube is in water of refractive index 1.33.

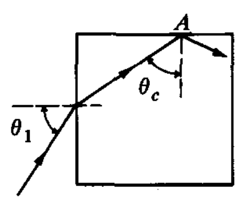


Figure 30-20

20. At what angle, with respect to the vertical, must a fish look in order to see a fisherman sitting on a distant shore? Assume that $n = 1.33$.

21. Generalize (30-15) to the case that the second medium is not air but has an index of refraction n_2 and show explicitly that here

$$\sin \theta_c = \frac{n_2}{n_1}$$

22. Two parallel light beams are incident from air onto the large face of a 45°-45°-90° prism of refractive index 1.30; see Figure 30-21. Calculate the angle between the refracted light beams when they emerge from the prism.

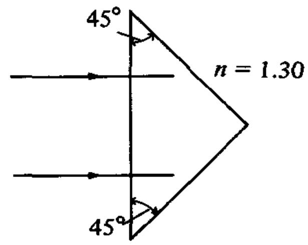


Figure 30-21

23. A light beam is incident normally from air onto a glass plate of refractive index 1.5. What fraction of the incident energy (expressed as a percent) enters the plate? What happens to the remainder?

24. Show explicitly, by use of (30-18), that the sum of the reflected and refracted intensities is numerically equal to the incident intensity.

25. Suppose that a light beam is incident from air perpendicularly onto the face of a certain transparent material. Calculate its refractive index if 50 percent of the light is reflected.

26. Show explicitly that for near-normal incidence ($\theta_1 \cong n_2\theta_2/n_1 \cong 0$), both (30-19) and (30-20) reduce to (30-17).

27. (a) Assuming near-normal incidence in Figure 30-15, show that the requirement that the tangential components of the electric field be continuous implies that

$$E_0 - E_0'' = E_0'$$

(b) Similarly, assuming that the tangential components of $\mathbf{H}$ ($\equiv \mathbf{B}/\mu_0$ since we assume that the media are not magnetic) are continuous, show that at normal incidence

$$n_1E_0 + n_1E_0'' = n_2E_0'$$

(c) By combining the results of (a) and (b) confirm the validity of (30-17).

***28.** Using the method of Problem 27, establish the validity of (30-19). Make use of the vectors in Figure 30-15 but do not assume the angles θ_1 and θ_2 to be small.

29. Show that the amplitudes in both (30-19) and (30-20) satisfy the relation

$$n_1 \cos \theta_1 E_0^2 = n_1 \cos \theta_1 E_0''^2 + n_2 \cos \theta_2 E_0'^2$$

What is the physical significance of this relation?

***30.** Consider, in Figure 30-22, a very narrow, rectangular path *ABCDA*, which straddles the interface between two media of refractive indices n_1 and n_2. Assume in the following that the length Δl is small but finite, and that the width of the rectangle tends to zero.

(a) Show that the tangential components of the electric field on the two sides of the interface must be the same by applying Faraday's law to the closed path *ABCDA*. (See Problem 34, Chapter 27.)

(b) Similarly, by applying another of Maxwell's equations to the path *ABCDA* show that the tangential components of **H** ($\equiv \mathbf{B}/\mu_0$ in the nonmagnetic case) are continuous. (See Problem 39, Chapter 28.)

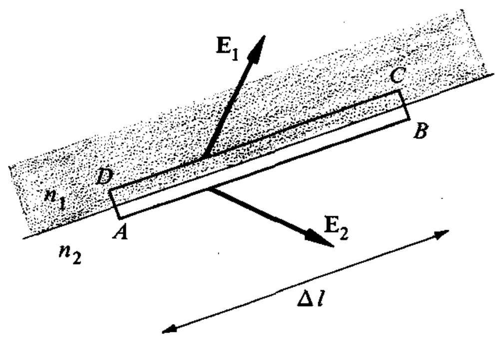


Figure 30-22

***31.** Using methods analogous to those in Problem 30, show by use of Gauss' law that the normal components of **D** ($\equiv \kappa\epsilon_0\mathbf{E}$) are continuous across the interface between two media. (*Hint*: Evaluate Gauss' law for a very short pillbox, which straddles the interface. See also Problem 38, Chapter 28.)

31 Mirrors and lenses

The beholding of the light is itself a more excellent and fairer thing than all the uses of it.

FRANCIS BACON

31-1 Introduction

In Chapter 30 we studied the laws which govern the reflection and the refraction of a collimated beam of light rays at the interface between two media. We shall now extend these studies to the more usual and complex situations involving the simultaneous reflections and refractions of large collections of such light beams.

Except under controlled laboratory conditions, the light waves emitted by a source normally consist of a superposition of many light beams, which go radially out in all directions from the source. It is for this reason, for example, that when an object—such as this page—is illuminated by a light source, it can be seen from many vantage points. The incident light is diffusely scattered so that each point of the object becomes, in effect, a point source from which light rays go out in all directions. Our perception of the object is determined by the particular subset of these rays which enters the eyes. Figure 31-1a, for example, shows some of the light rays that emanate from a point source S. Note that the light rays which enter the eyes of the observers O_1 and O_2 are distinct, so that each observer perceives S from his

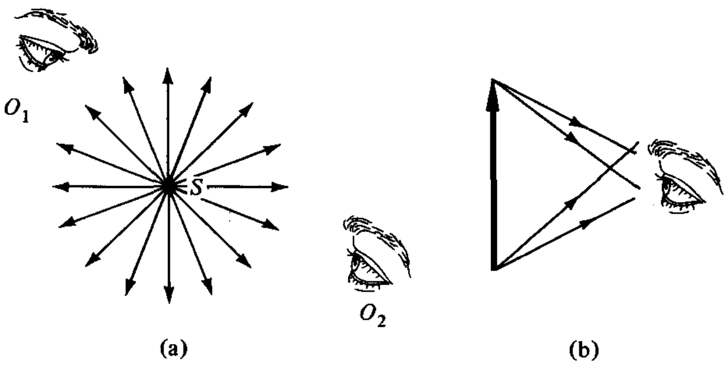


Figure 31-1

own point of view. Similarly, Figure 31-1b shows some of the light rays seen by an observer viewing an illuminated object. Although only the rays that go to the observer's eyes from the top and bottom of the arrow are shown, it is understood in such diagrams that light rays from intermediate points of the arrow enter his eyes as well. In other words, the observer perceives all points of the object, but for the sake of diagrammatic simplicity not all of the rays involved have been drawn.

Consider, in Figure 31-2, a point source S in air at a certain distance from the flat surface of a refracting medium. Each ray from S will, upon striking the interface, give rise to both a reflected and a refracted ray whose directions may be calculated from the laws of geometrical optics. An important question that arises in this connection is whether in addition to this predictable behavior of the individual light beams there are any significant regularities displayed by the *totality* of these reflected and refracted rays. The main purpose of this chapter is to establish an affirmative answer to this question by describing some of the quantitative laws that govern the behavior of such *collections* of light rays.

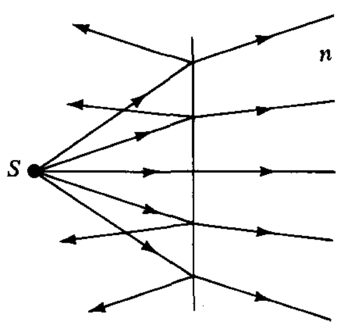


Figure 31-2

31-2 Plane mirrors

A *mirror* is a highly polished surface having the property that most of the light incident on it is reflected, and very little is therefore transmitted through it. Ordinary household or laboratory mirrors can be made by

evaporating a thin, metallic film on a smooth glass or Pyrex surface. Such mirrors are nearly 100 percent reflecting. Although we shall confine ourselves here to an analysis of reflections by such a mirror, it should be kept in mind that the results obtained are also applicable to ordinary dielectric surfaces, with the only exception that for the latter *not* all of the incident light is reflected.

Consider, in Figure 31-3, a point object O at a perpendicular distance s from a plane mirror MM'. As shown, some of the light rays that go radially outward from O will strike the mirror and be reflected backward in accordance with the law of reflection. An observer, such as O_1, looking into the mirror will see the object by virtue of some of these reflected rays which enter his eyes. To him, the object will *appear* to be at an *image point I*, which is at a certain distance s' *behind* the mirror. Most interesting is the fact that to *all* observers looking into the mirror the object appears to be at the same point I, which is thus known as the *image* of the object. This point, which lies along the perpendicular from O to the mirror, is precisely as far behind the mirror as the object is in front. Hence the object distance s and the image distance s' in the figure are equal.

An image of a point object of this type whose rays do not actually intersect in a point, but only appear to do so, is said to be a *virtual image*. It follows by reference to Figure 31-3 that all images formed in plane mirrors must be virtual. By contrast, for a concave spherical mirror, as will be seen in Section 31-3, it is very much possible for the light rays from a point object to intersect in a real point after reflection in the mirror. This type of an image is known as a *real image*. Note that for a real image there is an actual flow of energy through the image point, whereas no such flow is associated with a virtual image.

To establish the equality of s and s' consider, in Figure 31-4, an arbitrary ray OB from the object, which makes an angle θ with the ray OA perpendicular to the mirror. On striking the mirror, the ray OA is reflected back on itself, whereas the ray OB is reflected along the direction BC, which, according to the laws of geometrical optics, also makes an angle θ

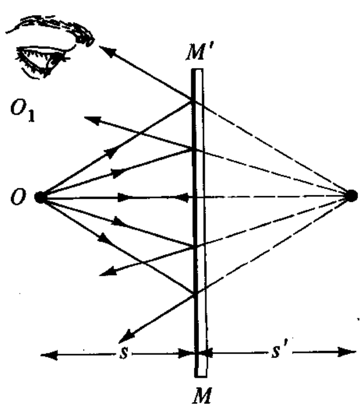


Figure 31-3

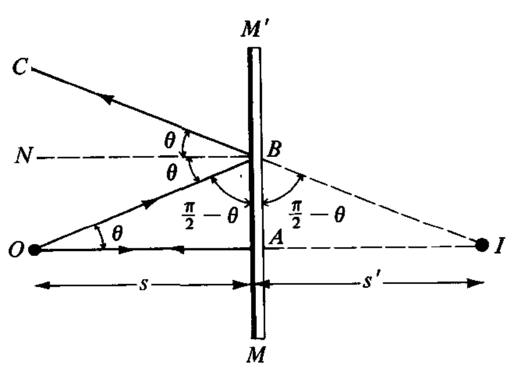


Figure 31-4

with the normal NB to the mirror. Let us extend the reflected rays AO and BC backward through the mirror and let them intersect at a point I at a certain distance s' behind the mirror. The two triangles OAB and ABI have the common side AB and by construction $\angle OAB = \angle IAB = 90°$ and $\angle OBA = \angle IBA = 90° - \theta$. Hence these two triangles are congruent, and therefore

$$s = s'$$

That is, regardless of the value for θ, the backward extension of the reflected ray BC intersects the corresponding extension of the perpendicular ray AO at the point I, which is precisely as far behind the mirror as the object is in front. Thus, the totality of all reflected rays appears to come from this image point I, and to *any* observer who sees only the reflected light the object appears to be at the image point I.

Figure 31-5 illustrates image formation of an object of finite dimensions in a plane mirror. For each point o of the object which is at a perpendicular distance s from the mirror there is a corresponding image point i the same distance behind it. Hence as shown in the figure by the dotted arrow, the image is precisely of the same size as the object but tilted by the angle α in the opposite sense.

Figure 31-6 shows the reflection of a complete three-dimensional object in a plane mirror. Again for each object point o in front of the mirror there is an image point i an equal distance behind it, with the line through o and i perpendicular to the mirror. Note that the sense of the z-axis of the image is opposite to that of the object, but that the senses of the x- and y-axes are unaltered. We describe this by saying that the *sense of front and back* for the image is opposite to that for the object. If you look into a mirror and move your right hand, then, consistent with the above arguments, you will find that the *left hand* of the image is the one that moves. Accordingly, we also say that the sense of left and right for the image in a plane mirror is opposite to that of the object.

Figure 31-7 illustrates the problem of image formation for an object O, which lies between two mirrors at right angles to each other. This time, as illustrated by the reflected rays, there will be three images: I_1, I_2, and I_3. Note that for I_2 the ray from O undergoes *two* reflections before entering

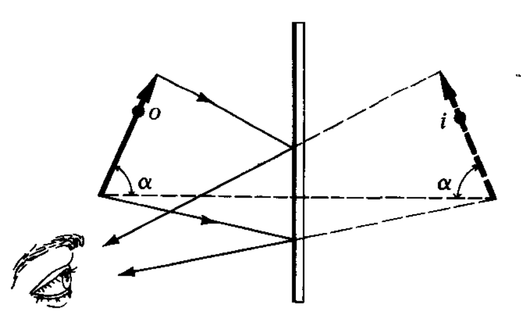


Figure 31-5

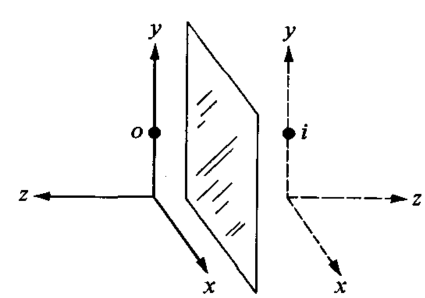


Figure 31-6

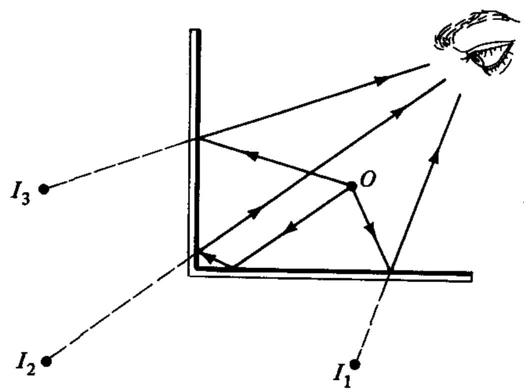


Figure 31-7

the observer's eye, so the sense of front and back for I_2 is the same as that for the object. The rays for the images I_1 and I_3 involve only one reflection, and thus for these the sense of front and back is opposite to that for the object. These features can be readily confirmed by looking into two mirrors at right angles. If you raise your right hand, you will find that each of the images on the left and the right will raise the left hand, while for the middle image the right hand will be raised.

31-3 Spherical mirrors

Besides the plane mirror, there are a number of other mirror geometries which are of physical and practical interest. One of these is the spherical mirror, and is the subject of this section and the next. We shall find that under certain conditions image formation also can take place in spherical mirrors.

Consider in Figure 31-8 an object O at a distance s from a *concave* spherical mirror of radius R. Let the point C represent the *center of curvature*—that is, the center of the spherical surface of which the mirror is a part—and define the line through O and C as the *axis* of the mirror. The intersection of this axis with the mirror defines its vertex V and, for reasons to be given below, we define the *focal point* F of the mirror as the point midway between V and C. Thus the focal point F is at a distance $R/2$ from both the vertex and the center of curvature.

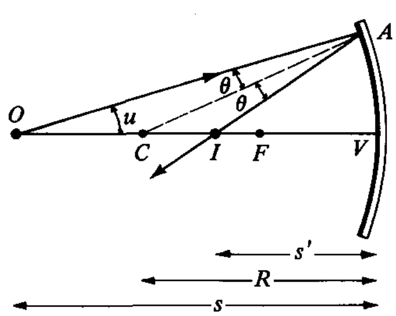


Figure 31-8

Figure 31-9 illustrates how the definitions of the various quantities above may be extended to situations other than that of an object at a distance greater than R from a concave spherical mirror. Figure 31-9a shows an object at a distance $s < R$ from a concave mirror of radius R. The above definition for the mirror axis, the vertex, and focal point are directly applicable to this case. The main difference is that now the object lies between the vertex and the center of curvature of the mirror. Correspondingly, Figure 31-9b shows how these definitions can be extended to the case of a *convex* mirror. The important difference for this case is that both the center of curvature C and the focal point F are here *behind* the mirror. The axis of the mirror is still defined to be the line joining O and C.

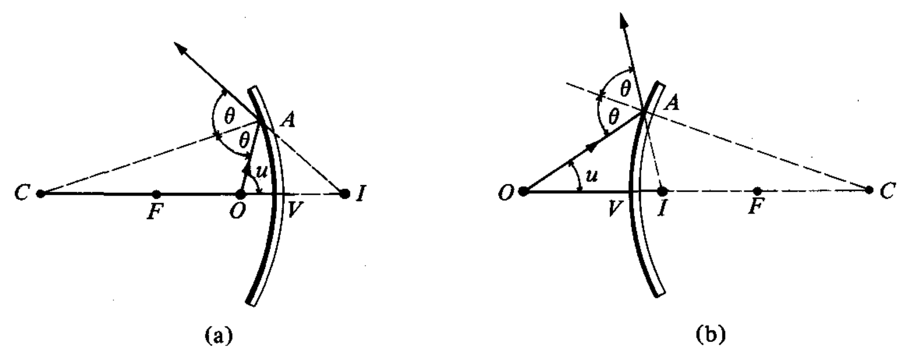


Figure 31-9

Let us now consider any one of the situations in Figures 31-8 and 31-9, and follow a typical ray OA from O, which makes an angle u with the axis and strikes the mirror at a point A. Since the radius vector CA is perpendicular to the mirror, it follows from the laws of geometrical optics that CA will bisect the angle between the incident and the reflected rays. The ray OV is itself perpendicular to the mirror and is thus reflected back on itself. It follows that if an image is formed in the spherical mirror, it must be at the point of intersection of the mirror axis OV and the reflected ray AI or its backward extension. For the situations in Figure 31-9 this image point I is analogous to that for a plane mirror and is *behind* the mirror; that is, the light rays are reflected by the mirror and *appear* to the observer to originate at the point I *behind* the mirror. These images are thus *virtual*. By contrast, for the situation in Figure 31-8 the image is *real*. Here the reflected rays go through the image point I, which is on the same side of the vertex as is the object. Hence for this case the observed light rays actually come from the image point I, and there is, in general, a flow of energy through the image point. By contrast, although energy is associated with all light rays, for a virtual image no energy actually goes through the image point itself; it only appears to do so.

So far in this discussion it has been implicitly assumed that image formation—whether it be real or virtual—actually takes place. We shall now establish that for *paraxial rays*, that is, rays for which the angle u in Figures

31-8 and 31-9 is sufficiently small that $u \cong \tan u \cong \sin u$, an image is actually formed in each case. Moreover, it will be shown that for a concave mirror the object distance s and the image distance s' are related by

$$\frac{1}{s}+\frac{1}{s'}=\frac{2}{R} \qquad \text{(concave)} \tag{31-1}$$

while for a convex mirror the corresponding formula is

$$\frac{1}{s}+\frac{1}{s'}=-\frac{2}{R} \qquad \text{(convex)} \tag{31-2}$$

where a negative value for s' means that the image is virtual and thus located a distance $|s'|$ behind the mirror.

Alternatively, we may express (31-1) and (31-2) in the form

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f} \tag{31-3}$$

where f is the *focal length* of the mirror and has the value $+R/2$ for a concave mirror and $-R/2$ for a convex one. Physically, f is the image distance associated with an object at infinity. As illustrated in Figure 31-10a, for the case of a concave mirror the incoming parallel rays from a very distant object are brought together at the focal point F of the mirror. Correspondingly, as shown in Figure 31-10b, the reflected rays from a point object at the focal point of a concave mirror are parallel to the mirror axis. The analogous result for a convex mirror is shown in Figure 31-13 and will be considered in Section 31-4.

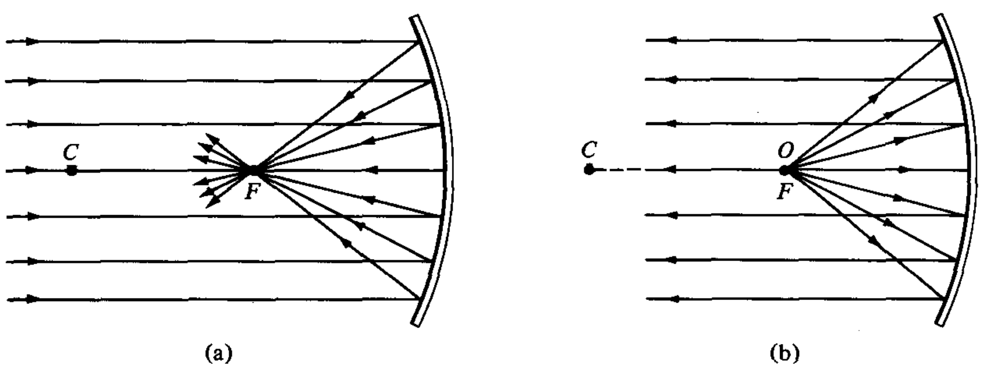


Figure 31-10

To establish (31-1) suppose that $s > R/2$, and consider in Figure 31-11 an arbitrary ray OA from the object, which makes an angle u with the axis. Define the angles α, β, and θ as in the figure, and let h be the length of the perpendicular from A to the axis and ϵ the distance from its foot to the vertex V. Since the exterior angle of a triangle is the sum of the remote interior angles it follows that

$$\beta = \alpha + \theta \qquad \alpha = u + \theta$$

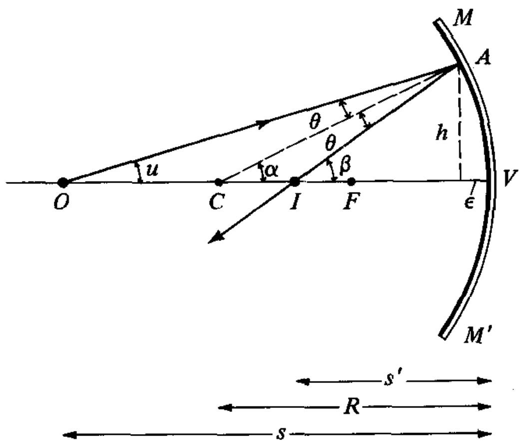


Figure 31-11

Eliminating the angle θ, we obtain

$$u + \beta = 2\alpha \tag{31-4}$$

Moreover, reference to the figure shows that

$$\begin{aligned} \tan u &= \frac{h}{s - \epsilon} \\ \tan \beta &= \frac{h}{s' - \epsilon} \\ \tan \alpha &= \frac{h}{R - \epsilon} \\ \epsilon &= R(1 - \cos \alpha) \end{aligned} \tag{31-5}$$

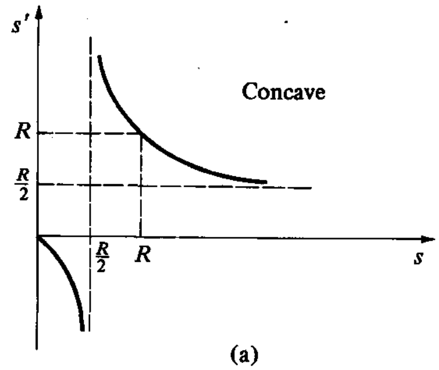
since $s = \overline{OV}$, $R = \overline{CV}$, and $s' = \overline{IV}$. Now for fixed s and R, (31-4) and (31-5) represent only five relations among the six quantities u, β, α, h, ϵ, s'. Hence they do *not* associate a unique image distance s' for *all* angles u. However, and this is a most important point, the paraxial rays emanating from O do form a unique image for any object distance s. For these rays, with which we shall concern ourselves exclusively from now on, (31-5) may be approximated by

$$u = \frac{h}{s} \qquad \beta = \frac{h}{s'} \qquad \alpha = \frac{h}{R} \qquad \epsilon = \frac{R\alpha^2}{2} \cong 0 \tag{31-6}$$

since if u is small, so that $u \approx \tan u \approx \sin u$, then so are each of the remaining angles α, β, and θ. Substituting the first three relations of (31-6) into (31-4), we find, after canceling the common factor h, the desired relation in (31-1).

The validity of (31-1) for $s < R/2$ as well as (31-2) can be established by analogous arguments. Details will be found in the problems.

Figure 31-12a shows a plot of the image distance s' as a function of s for a concave mirror, as predicted by (31-1). Note that if $s > R/2$, then $s' > 0$, and the image is real, whereas if $s < R/2$, then $s' < 0$ and, as confirmed in Figure



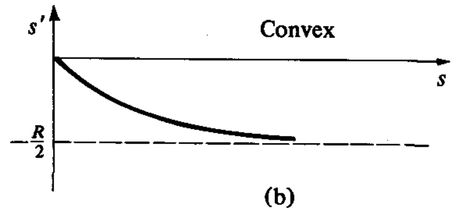


Figure 31-12

31-9a, the image is virtual. If the object is at the focal point F, so that $s = R/2$, then according to the graph $s' = \infty$. Hence, as shown in Figure 31-10b, the reflected rays are parallel to the mirror axis in this case. Figure 31-12b shows the corresponding plot of s' against s for a convex mirror. As can be confirmed by reference to Figure 31-9b, the image is virtual for any object distance in this case. Hence, as shown in the graph, only negative values for s' occur.

31-4 Applications

In this section we apply (31-1) and (31-2) to several concrete physical situations.

Example 31-1 A point object is at a distance of 20 cm from a spherical mirror of radius 10 cm. What is the nature and location of the image if the mirror is (a) Concave? (b) Convex?

Solution

(a) Here we are given the values $s = 20$ cm and $R = 10$ cm. Substitution into (31-1) yields

$$s' = \frac{20}{3}\text{ cm} = 6.7\text{ cm}$$

Since $s' > 0$, the image is real and on the same side of the mirror as the object.

(b) For the convex mirror, we must use (31-2). Substituting the value $s = 20$ cm, we find that

$$s' = -4.0\text{ cm}$$

so the image is virtual this time.

Example 31-2 Parallel rays are incident on a spherical mirror having a radius of 0.5 meter. What is the nature and location of the image formed by these rays if the mirror is: (a) Concave? (b) Convex?

Solution

(a) Substituting the given values, $s = \infty$ and $R = 0.5$ meter, into (31-1), we find that

$$\frac{1}{\infty} + \frac{1}{s'} = \frac{2}{0.5 \text{ m}}$$

that is,

$$s' = 0.25 \text{ m}$$

The image is real and, as shown in Figure 31-10a, it is located at the focal point of the mirror at a distance $R/2$ from the vertex.

(b) Substituting $s = \infty$ and $R = 0.5$ meter into the convex mirror formula in (31-2), we find in a similar way that

$$s' = -0.25 \text{ m}$$

and this is shown in Figure 31-13. Note that the image is virtual and located a focal length ($R/2$) *behind* the mirror.

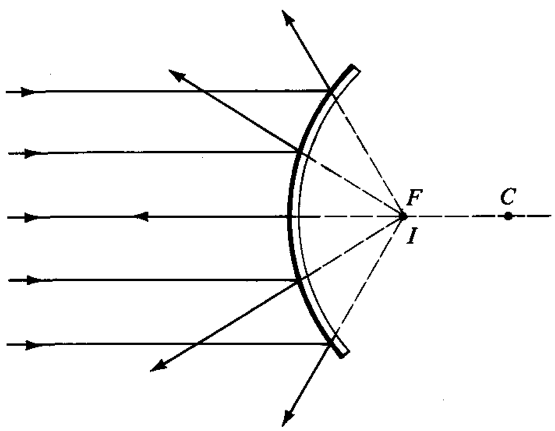


Figure 31-13

Example 31-3 A point object O is at a distance $(3/4)R$ from a concave mirror of radius R; see Figure 31-14.

(a) What is the nature and location of the image?

(b) How can an observer tell which of O and I is the object and which is the image?

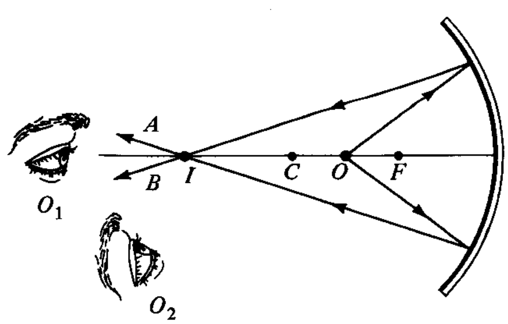


Figure 31-14

Solution

(a) The substitution of the given values for s and R into (31-1) yields

$$\frac{1}{(3/4)R} + \frac{1}{s'} = \frac{2}{R}$$

and thus

$$s' = +\frac{3}{2}R$$

so the image is real.

(b) A person viewing the object O directly will see it regardless of which observation point he happens to choose. This follows since a point object is one that radiates light in all directions. On the other hand, because of the finite extent of the mirror, the reflected rays are confined—as shown in Figure 31-14—to the interior of the cone AIB. In other words, the observer O_1 will see the image, since some of the reflected rays will enter his eyes. On the other hand, an observer O_2 looking at this same point I in space but from outside of the cone AIB will see nothing, since the reflected rays that would enter his eyes if the mirror were large enough simply are not reflected. Thus by changing points of view we can tell the object and the image apart. That is, there are always positions of the observer for which the image is invisible!

31-5 Finite objects

Consider, in Figure 31-15, an object of finite size—say an arrow of length y—at a distance s from and perpendicular to the axis of a concave spherical mirror of radius R. The image of the point A at the base of the arrow will occur at the point A', which is on the axis and at a distance s' from the vertex as determined by (31-1). The object distance of the point B, which represents the head of the arrow, is slightly greater than is the object distance s for the point A, but for simplicity let us neglect this difference by confining ourselves to small objects. In general, if the object is not small, the image will undergo a distortion known as *curvature of field.* As is implied in the figure, we shall ignore these and other *spherical aberrations*, as they are

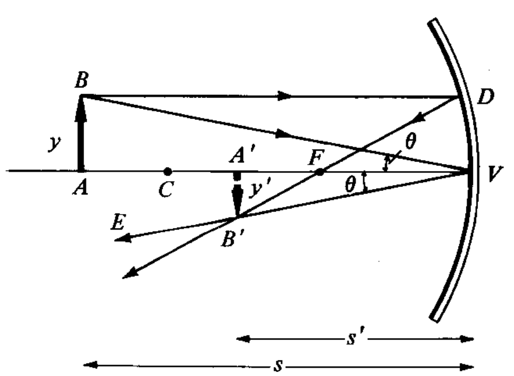


Figure 31-15

referred to, so that the image, $A'B'$, will always be assumed to be parallel to the object, AB.

The position of the image B' of the point B may be found by drawing any two of the following four rays. The two of these that have been drawn in the figure are (1) the ray BD parallel to the mirror axis, which after reflection goes through the focal point F; and (2) the ray BV to the vertex of the mirror, whose reflection VE makes the same angle θ with the axis that the incident ray does. The remaining two rays, which have not been included in the figure, are (3) the ray BC through the center of curvature, which will be reflected back on itself; and (4) the ray BF through the focal point, which after reflection will be parallel to the axis. (See Figure 31-16.) The intersection of any two of these reflected rays will determine a unique point B' at a certain distance y' below the axis. The perpendicular from B' to the axis will intersect the axis at the point A', which is the image of the point A. All other rays drawn from points on the arrow between A and B will be imaged between points A' and B', and thus there is no need to draw any additional rays.

Consider the two right triangles ABV and $VA'B'$. They are similar to each other since their angles are equal in pairs. Hence we obtain

$$\frac{y}{y'} = \frac{s}{s'} \tag{31-7}$$

By use of this formula and (31-1) the size of the image y' for any given object can be readily determined.

In connection with a discussion of the reflection of objects of finite size in a spherical mirror it is convenient to define the *lateral magnification*, m, of the object to be the ratio of the size of the image to that of the object. The sign of m is by definition negative if the image is inverted relative to the object, and positive otherwise. For the situation in Figure 31-15 with a concave mirror, this means that

$$m = -\frac{y'}{y} \tag{31-8}$$

where the minus sign is required since in this case the image is inverted. Substitution for the ratio y'/y from (31-7) leads to

$$m = -\frac{s'}{s} \tag{31-9}$$

which together with (31-1) and (31-2) represent the basic tools necessary for the analysis of image formation in spherical mirrors.

Although (31-9) has been derived only for the special situation shown in Figure 31–15—that is, where the object is at a distance $s > R$ from a concave spherical mirror and for which the image is real and inverted—it is true in general. That is, regardless of whether the mirror is concave or convex or whether the image is real or virtual or whether it is erect or inverted, (31-9) is invariably correct. The validity of this feature is illustrated in Figure 31-16

for various special cases. Figure 31-16a shows an object AB at a point between the focal point F and the center of curvature C. Its image is found by drawing the ray BD parallel to the axis, which after reflection goes through the focal point, and the ray BF through the focal point F, which after reflection is parallel to the axis. The intersection of these two rays as well as the ray CB through the center of curvature determines the image point B'. Note that consistent with (31-9) the image is inverted and larger than the object. Similarly, in Figure 31-16b, where the object distance is less than the focal distance, we see that the image is virtual and erect. This means that the image distance s' is negative and thus, consistent with (31-9), the magnification is positive. Finally, Figure 31-16c shows the case of an object in front of a convex mirror. The image is virtual, thus $s' < 0$; and therefore, consistent with (31-9), the lateral magnification m is positive. Hence the image is erect, as shown.

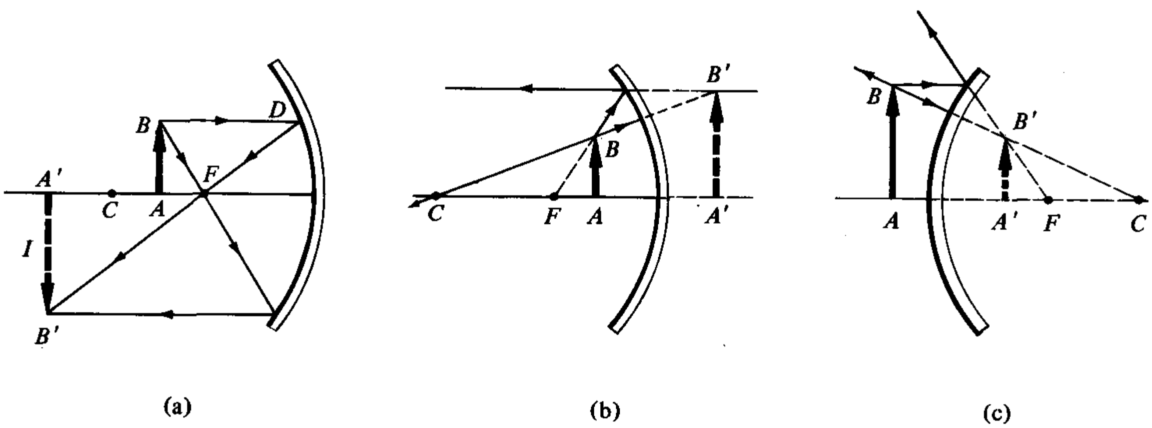


Figure 31-16

Example 31-4 An object of height 2.0 cm is perpendicular to the axis and at a distance of 5.0 cm from a concave spherical mirror of radius 15 cm.

(a) Locate and describe the nature of the image.
(b) Calculate the lateral magnification.
(c) What is the size of the image?

Solution

(a) Substituting the given data, $R = 15$ cm and $s = 5.0$ cm, into (31-1) yields

$$\frac{1}{5\text{ cm}} + \frac{1}{s'} = \frac{2}{15\text{ cm}}$$

so that $s' = -15$ cm. Hence the image is virtual and at a distance of 15 cm behind the mirror.

(b) Substituting the above values for s and s' into (31-9) yields for m

$$m = -\frac{s'}{s} = -\frac{-15\text{ cm}}{5\text{ cm}}$$
$$= +3$$

Since this is positive, it follows that the image is erect.

(c) Since $m = +3$, it follows by use of (31-8) that the image is three times as large as the object. Since the object has a height of 2 cm, the height of the image is 6 cm.

Example 31-5 An object is at a distance s from a spherical mirror of radius R. If the object is displaced by a small amount ds, show that the corresponding displacement ds' of the image is

$$\frac{ds'}{ds} = -m^2$$

where m is the lateral magnification, and describe this result qualitatively.

Solution Let us start with the basic relation

$$\frac{1}{s} + \frac{1}{s'} = \pm\frac{2}{R}$$

where the upper sign is for a concave mirror and the lower for a convex one. Taking differentials on both sides of this relation we find, since R is constant, that

$$-\frac{ds}{s^2} - \frac{ds'}{s'^2} = 0$$

Solving for the ratio ds'/ds, we thus obtain

$$\frac{ds'}{ds} = -\left(\frac{s'}{s}\right)^2 = -m^2$$

where the last equality follows by (31-9). The quantity ds'/ds is known as the *longitudinal magnification.*

Even though the lateral magnification m can be either positive or negative, the ratio ds'/ds is negative in all cases. This means that if an object is brought closer to a mirror—that is, if $ds < 0$—then $ds' > 0$, so that if the image is real, then it recedes from the mirror while if it is virtual it approaches the mirror. Contrariwise, if $ds > 0$, so that the object recedes from the mirror, then $ds' < 0$, so that the image either approaches or recedes from the mirror depending on whether the image is real or virtual.

31-6 Image formation by spherical refracting surfaces

We now turn from the problem of image formation in mirrors, and for the remainder of this chapter consider the related problem of the formation of images by light rays refracted at the spherical interface between two media.

Figure 31-17 shows a point object O at a distance s_1 from the spherical interface of radius R between two media of refractive indices n_1 and n_2, respectively. The refracting surface is assumed to be convex in this case. The perpendicular ray OC from the object to the center of curvature C cuts the surface at the vertex V and defines the *axis* of the system. A second ray OA, which makes an arbitrary angle u with respect to the axis, has also been

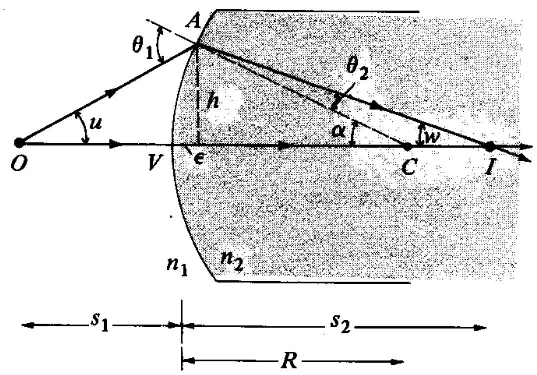


Figure 31-17

drawn in the figure. Assuming that $n_2 > n_1$, this ray will be refracted toward the normal into the second medium and, as shown, it will intersect the axis at a point I at a certain distance s_2 from the vertex. The angle of incidence θ_1 and the angle of refraction θ_2 in the figure are defined with respect to the normal to the surface at A, which is the radius CA from the center of curvature C.

Just as for spherical mirrors, it will be established below that for *paraxial rays* image formation also takes place for spherical refracting surfaces. For the specific case shown in the figure, with the object in the n_1-medium on the convex side and at a distance s_1 from the vertex, an image will be formed at a distance s_2 from the vertex given by

$$\frac{n_1}{s_1}+\frac{n_2}{s_2}=\frac{n_2-n_1}{R} \qquad \text{(convex)} \qquad \textbf{(31-10)}$$

If s_2, is positive then the image is real and is formed in the n_2-medium to the right of the interface, as shown in Figure 31-17. On the other hand, if $s_2 < 0$, then, as shown in Figure 31-18, a virtual image is formed in the same medium as is the object. Correspondingly, for the case of the concave interface, as in Figure 31-19, the relation between the object and the image distance is

$$\frac{n_1}{s_1}+\frac{n_2}{s_2}=-\frac{n_2-n_1}{R} \qquad \text{(concave)} \qquad \textbf{(31-11)}$$

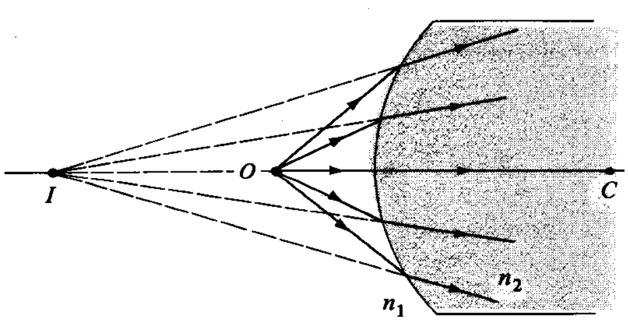


Figure 31-18

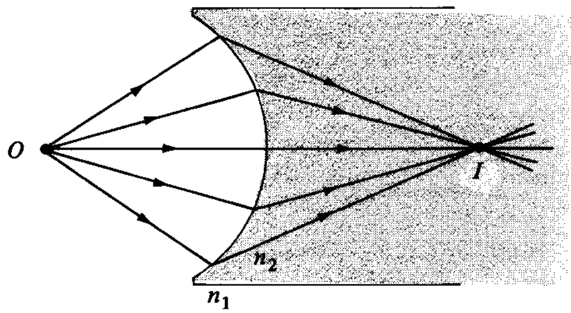


Figure 31-19

Note that, just as for the spherical mirror formulas in (31-1) and (31-2), the relations (31-10) and (31-11) may be obtained from each other by the substitution $R \rightleftarrows -R$. The use of these formulas makes possible the calculation of the image distance s_2 for any given object distance s_1. Positive values for s_2 imply a real image, as in Figures 31-17 and 31-19, and for these the image is on the opposite side of the interface from the object. Figure 31-18 shows the case of a virtual image, for which s_2 is negative, and here the image is in the same medium as is the object.

To derive (31-10), let us define the various angles u, θ_1, θ_2, α, and w, and the distances s_1, s_2, h, and ϵ as in Figure 31-17. Since the exterior angle of a triangle is the sum of the remote interior angles, it follows that

$$\alpha = w + \theta_2 \qquad \theta_1 = \alpha + u \tag{31-12}$$

while according to Snell's law,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{31-13}$$

Reference to the figure makes it possible to establish the additional relations

$$\begin{aligned} \tan u &= \frac{h}{s_1 + \epsilon} \\ \tan w &= \frac{h}{s_2 - \epsilon} \\ h &= R \sin \alpha \\ \epsilon &= R(1 - \cos \alpha) \end{aligned} \tag{31-14}$$

Again, as for the case of the spherical mirror, for fixed values of s_1, R, n_1, and n_2, these seven relations do not yield a unique value for the eight quantities α, θ_1, θ_2, u, w, h, ϵ, and s_2. Hence there does *not* exist a unique image distance s_2 for all values of u. However, for paraxial rays, (31-13) and (31-14) become

$$n_1\theta_1 = n_2\theta_2 \qquad u = \frac{h}{s_1} \qquad w = \frac{h}{s_2} \qquad h = R\alpha \qquad \epsilon \simeq 0 \tag{31-15}$$

and, substituting for θ_1 and θ_2 into the first of these by use of (31-12), we obtain

$$n_1(\alpha + u) = n_2(\alpha - w)$$

or

$$n_1 u + n_2 w = \alpha(n_2 - n_1)$$

Finally, eliminating the angles u, w, and α by use of (31-15), and canceling the common factor h, the sought-for relation in (31-10) results. The validity of (31-10) for a virtual image, as in Figure 31-18, can be established in the same way.

To derive (31-11) we may now proceed as follows. Reference to Figures 31-17 and 31-19 shows that if, for a given value for s_1, the object is placed at the image point (of a real image), then the resultant image occurs at the old object point. That is, if an object is placed at the point I in either of these figures, its image occurs at the point O. For the effect of exchanging the image and the object is to reverse the sense of direction of all light rays, and the laws of optics work just as well for the reversed rays as they do for the original ones. For the case of a mirror, this reversal means that $s' \rightleftarrows s$ and (31-1) and (31-2) are invariant under this exchange, as they must be. In the present case the interchange of the object and image means that $s_1 \rightleftarrows s_2$ and $n_1 \rightleftarrows n_2$, and making these substitutions in (31-10) we are led directly to (31-11). In other words, both (31-10) and (31-11) are invariant under the simultaneous replacements $s_1 \rightleftarrows s_2$ and $n_1 \rightleftarrows n_2$ and $R \to -R$. Therefore provided negative values for R are used for concave surfaces, (31-10) is very generally applicable to both concave and convex surfaces.

Graphs of (31-10) for the special case $n_2 > n_1$ are shown in Figure 31-20a and for $n_2 < n_1$ in Figure 31-20b. Note the similarity between these graphs and the corresponding ones for a mirror in Figure 31-12. For the case $n_1 < n_2$, if the object is placed a distance $Rn_1/(n_2 - n_1)$ from the vertex, the refracted rays are parallel to the axis and thus are imaged at infinity.

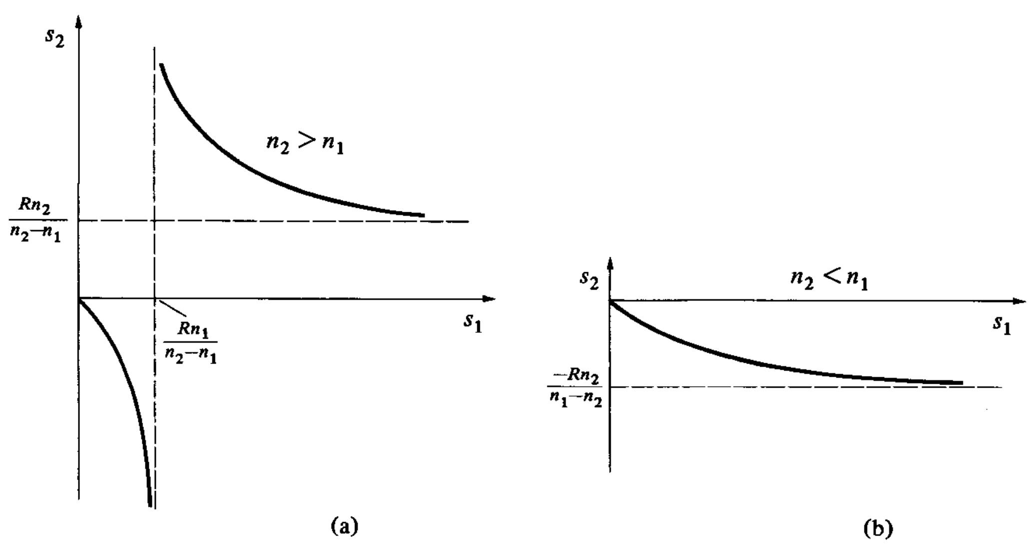


Figure 31-20

Example 31-6 A bundle of parallel rays are traveling in air and are incident from the left onto a spherical surface of glass ($n = 4/3$) of radius 10 cm. Describe the nature and location of the image if the surface is:

(a) Convex, as in Figure 31-17.
(b) Concave, as in Figure 31-19.

Solution

(a) The substitution of the given values $s_1 = \infty$, $R = +10$ cm, $n_1 = 1$, $n_2 = 4/3$ into (31-10) yields

$$\frac{1}{\infty} + \frac{4/3}{s_2} = \frac{(4/3) - 1}{10\text{ cm}}$$

that is,

$$s_2 = +40\text{ cm}$$

Hence the image is real and on the right-hand side of the interface.

(b) For this case, the parameter values are $s_1 = \infty$, $R = -10$ cm, $n_1 = 1$, and $n_2 = 4/3$. Substitution into (31-10) leads to

$$\frac{1}{\infty} + \frac{4/3}{s_2} = -\frac{(4/3) - 1}{10\text{ cm}}$$

and this yields

$$s_2 = -40\text{ cm}$$

Thus, after being refracted, the rays diverge and appear to come from the virtual image at a point 40 cm to the left of the interface.

Example 31-7 A point source of light is on the bottom of a tank of water ($n = 1.33$) of depth h. What is the apparent depth h' as seen by an observer who is in air and views the light at normal incidence?

Solution Since the observer views the source at normal incidence, only paraxial rays are involved and we may use (31-10). The surface of the water has zero curvature, so the parameter values are

$$R = \infty \qquad n_1 = \frac{4}{3} \qquad n_2 = 1 \qquad s_1 = h$$

Substitution into (31-10) yields

$$\frac{4/3}{h} + \frac{1}{h'} = -\frac{(4/3) - 1}{\infty}$$

and, consistent with the result in Example 30-4, this leads to

$$h' = -\frac{3}{4}h$$

31-7 The lensmaker's equation

Consider, in Figure 31-21, an object O on the *axis* of and at a distance s from a *thin lens* of refractive index n. Let R_1 and R_2 be the radii of curvature of its left and right faces, respectively, and, to be specific, suppose that their respective centers of curvature C_1 and C_2 are located as shown. The line joining C_1 and C_2 is called the *axis* of the lens. And the lens is "thin" provided that its maximum thickness is small compared to its two radii of curvature. In the following we shall be concerned only with thin lenses.

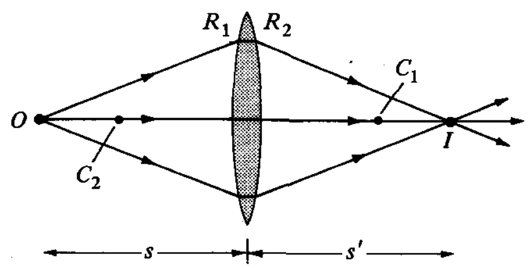


Figure 31-21

The main purpose of this section is to establish that the distance s' of the image I formed by a thin lens in air is related to the object distance s by

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f} \tag{31-16}$$

where f is the "focal length" of the lens and is given by the lensmaker's equation

$$\frac{1}{f}=(n-1)\left[\frac{1}{R_1}+\frac{1}{R_2}\right] \tag{31-17}$$

In utilizing this formula for the focal length f, the parameter R_1 is to be taken to be positive if, as in Figure 31-21, C_1 is to the right of the first refracting surface, and negative otherwise. Similarly, R_2 is positive if, as in the same figure, C_2 is to the *left* of the lens, and negative otherwise. Hence the focal length f can be positive or negative depending on the positions of the centers of curvature. The term *positive* or *converging lens* is used to characterize a lens with a positive focal length, and similarly a *negative* or a *diverging* lens is one whose focal length is negative. The particular lens shown in Figure 31-21 is known as a *double convex* lens and it has positive focal length. By contrast, the lens in Figure 31-22b is a *double concave* lens and since for it both R_1 and R_2 in (31-17) are negative, so is its focal length. Similarly the *plano-convex* and the *plano-concave* lenses in Figure 31-23 have positive and negative focal lengths, respectively.

The significance of the term "focal length" for the quantity represented by the symbol f in (31-16) and (31-17) can be seen by reference to Figure 31-22. Part (a) shows a beam of light rays incident on and parallel to the axis of a

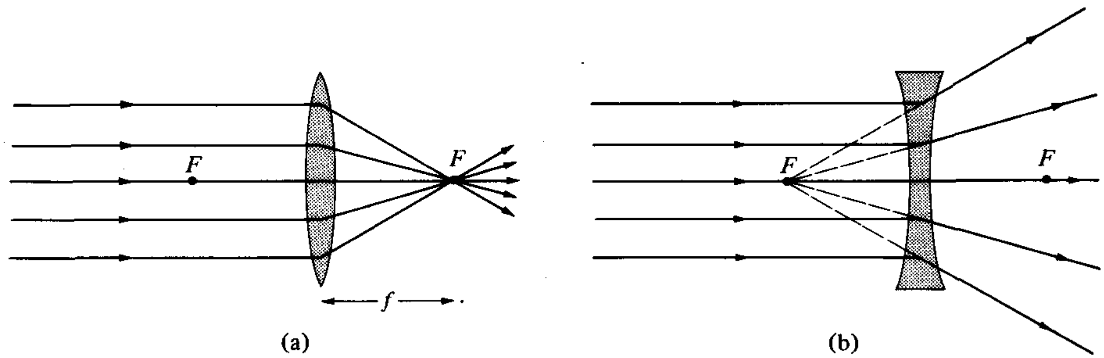


Figure 31-22

double convex lens, with its positive focal length. Since the object for these rays is at infinity, we find on setting $s = \infty$ in (31-16) that the image distance s' is equal to the focal length f. In other words, the focal length f is the distance from the positive lens of an image formed by an object at infinity. Equivalently, since an interchange of the object and its image can be achieved by reversing the direction of the light rays, the focal length is the distance from a positive lens at which a point source must be placed so that the emerging rays on the other side of the lens are parallel to the axis. Because of the inherent symmetry under the exchange $R_1 \rightleftarrows R_2$ in the lensmaker's equation in (31-17), the focal length f is the same on the two sides of the lens. Hence, as shown in the figure, the two focal points F of the lens are at the same distance f on opposite sides of the lens. Figure 31-22b shows a parallel beam incident on a double concave lens with its negative focal length. This time, setting $s = \infty$ in (31-16), we find that $s' = f$, but now f is negative! Therefore the image distance is negative and, as shown, the image is virtual and on the same side of the lens as is the object. For this case, the rays emerging from the lens diverge and thus appear to come from the side of the lens on which the rays are incident.

Comparison of (31-16) with the corresponding formulas for the spherical mirror in (31-1) and (31-2) shows that the graph in Figure 31-12 is also a graph of (31-16), provided that $R/2$ is interpreted as the focal length f of the lens. The important distinction is that for the mirror a positive image distance s' corresponds to a real image on the *same* side of the mirror as is the object.

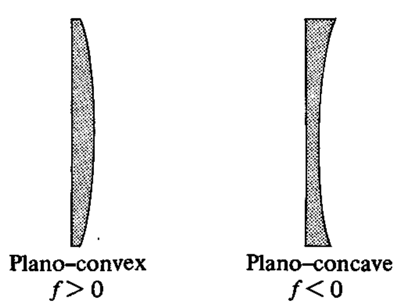


Figure 31-23

For a thin lens, on the other hand, a positive image distance means that the image is on the *opposite* side of the lens from the object.

To establish the validity of (31-16) and (31-17), let us consider the situation in Figure 31-21 and apply (31-10) and (31-11) at the two surfaces. For the refraction of the light rays from O at the first surface, the appropriate parameter values to substitute into (31-10) are $n_1 = 1$, $n_2 = n$, $s_1 = s$, $s_2 = s_2$, and $R = R_1$. Hence

$$\frac{1}{s} + \frac{n}{s_2} = \frac{n-1}{R_1} \tag{31-18}$$

Since the lens is presumed thin, it follows that the object distance for the refraction of these rays at the second surface is $-s_2$. Hence the appropriate parameter values to be used in (31-11) for the refraction of the light rays at the second surface are $s_1 = -s_2$, $s_2 = s'$, $n_1 = n$, $n_2 = 1$, and $R = R_2$. Hence

$$\frac{n}{-s_2} + \frac{1}{s'} = -\frac{(1-n)}{R_2} \tag{31-19}$$

Finally, eliminating s_2 between (31-18) and (31-19) and noting the formula for f in (31-17), we obtain the basic relation in (31-16).

Example 31-8 Assuming that the curved surface of a plano-convex lens (Figure 31-23) has a radius of curvature of 10 cm and that the refractive index of glass is 1.5, calculate the focal length of the lens.

Solution Assuming to be specific that the object is on the flat side of the lens, the radii of the lens are $R_1 = \infty$ and $R_2 = +10$ cm. Substituting these values into the lensmaker's equation, we obtain

$$\begin{aligned}\frac{1}{f} &= (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)\\ &= (1.5-1)\left(\frac{1}{\infty} + \frac{1}{10\text{ cm}}\right)\\ &= 0.05/\text{cm}\end{aligned}$$

whence

$$f = +20\text{ cm}$$

The positive sign means that the lens is converging, so that, for example, a beam of parallel rays will be focused to a real image on the opposite side of the lens and at a distance 20 cm from it.

Example 31-9 Suppose that for the double concave lens in Figure 31-22b the two radii of curvature are -10 cm and -20 cm and that the refractive index of glass is 1.5. Calculate the focal length of this lens.

Solution This time the values for both R_1 and R_2 are negative. Substitution into the

lensmaker's equation leads to

$$\frac{1}{f}=(n-1)\left(\frac{1}{R_1}+\frac{1}{R_2}\right)$$

$$=(1.5-1)\left(-\frac{1}{10\text{ cm}}-\frac{1}{20\text{ cm}}\right)$$

$$=-\frac{3}{40\text{ cm}}$$

Hence

$$f=-13.3\text{ cm}$$

Example 31-10 A parallel beam of light rays is incident on a glass sphere of radius 6 cm and of refractive index 1.5. Calculate the value of y where the rays come to a focus; see Figure 31-24.

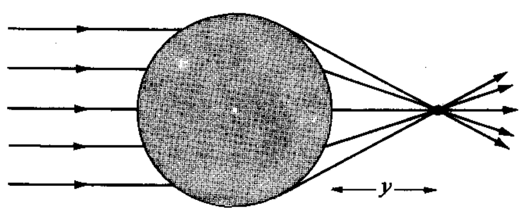


Figure 31-24

Solution Since the lens is *not thin*, (31-16) and (31-17) are *not* applicable. Let us therefore proceed as in the derivation of these formulas themselves.

At the first surface, the appropriate parameter values are $s_1=\infty$, $n_1=1$, $n_2=1.5$, and $R=6$ cm. Substituting these into (31-10), we find that

$$0+\frac{1.5}{s_2}=\frac{1.5-1}{6\text{ cm}}$$

so that $s_2=18$ cm. At the second surface, the corresponding parameter values are $s_1=12\text{ cm}-18\text{ cm}=-6$ cm, $s_2=y$, $n_1=1.5$, $n_2=1$, and $R=-6$ cm since the distance between the curved surfaces is 12 cm. A second application of (31-10) thus yields

$$-\frac{1.5}{6\text{ cm}}+\frac{1}{y}=\frac{(1-1.5)}{-6\text{ cm}}$$

Solving for y we obtain

$$y=3\text{ cm}$$

31-8 Images formed by thin lenses

Consider, in Figure 31-25, an object AB, of height y, perpendicular to the axis of and at a distance s from a thin, converging lens of focal length f. Assuming y to be sufficiently small so that any optical distortions can be

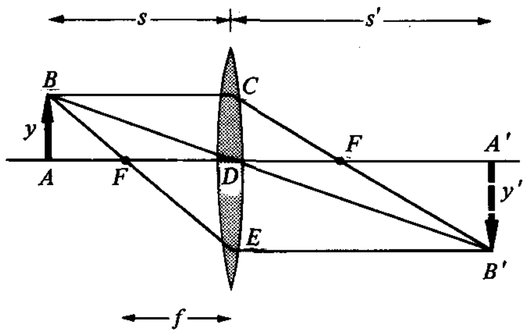


Figure 31-25

neglected, the image $A'B'$ of the object will be formed at a certain distance s' from the lens in accordance with (31-16). The lateral magnification m of the image is defined by

$$m = -\frac{y'}{y} \tag{31-20}$$

where y' is the size of the image. The purpose of the minus sign in this relation is so that m will be negative if, as in Figure 31-25, the image is inverted, and positive if the image is erect. This definition is the same as that for a spherical mirror in (31-8).

To determine the height y' of the image, it is necessary to find the image B' of the object point B. (The location of the image A' of the base of the object can be found directly from (31-16).) Figure 31-25 shows three rays, BCB', BDB', and BEB', the intersection of any two of which determine the image point B'. The ray BC is drawn parallel to the axis of the lens and thus, after being refracted in the lens, it will go through the focal point F as shown. Correspondingly, if the ray BE is drawn through the focal point F on the same side of the lens as the object, then after refraction it will be parallel to the lens axis. Finally, the ray BD through the center of the lens will emerge on the other side without being deviated. This follows from the result of Problem 12 in Chapter 30 and its associated Figure 30-19. The fact that all three of these rays (as well as all others that originate at B) must come together at the single point B' may be used as a consistency check when making use of this graphical method—illustrated in the figure—to determine the image formed by a thin lens.

Proceeding with the calculation of the height y' of the image, we note that in Figure 31-25 the triangles ABD and $A'B'D$ are similar. For they are both right triangles with the two angles $\angle BDA$ and $\angle B'DA'$ vertical angles and thus equal to each other. Hence

$$\frac{s}{s'} = \frac{y}{y'}$$

and substituting this into (31-20) we find that the lateral magnification m becomes

$$m = -\frac{s'}{s} \tag{31-21}$$

Even though this relation has been derived only for the converging lens in Figure 31-25, its validity extends to all thin lenses, as may be verified by applying the above argument to the situations in Figures 31-26 and 31-27.

By contrast to the situation in Figure 31-25, for which the object distance s exceeds the focal length f of the lens, Figure 31-26 shows the ray diagram for the two other possibilities: $s < f$ and $s = f$. As shown in Figure 31-26a, if the object is within a focal length of the lens, the emerging rays diverge and thus the image is virtual and larger than the object. Also consistent with (31-21), since $s' < 0$ for a virtual image, m is positive and the image is erect as shown. When the object is at the focal point itself, according to (31-16) $s' = \infty$. Substitution into (31-21) thus leads to $m = -\infty$, and hence, as implied by the parallel rays on the right side of the lens in Figure 31-26b, the image is inverted and "infinitely" large in this case.

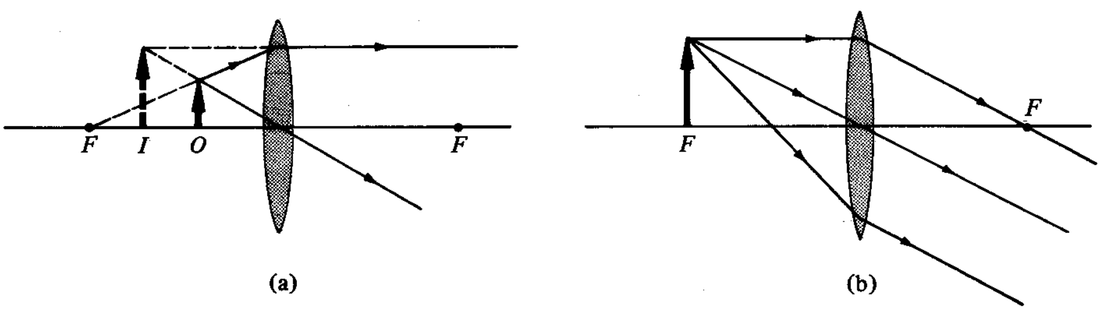


Figure 31-26

Figure 31-27 shows the analogous situation for the images formed by a thin, divergent lens with its negative focal length. Since only virtual images are possible for such a lens, the image distance s' is always negative. Hence, according to (31-21), m must be positive. As shown, therefore, for all three cases— $s > |f|$, $s < |f|$, and $s = |f|$—the image is virtual and erect.

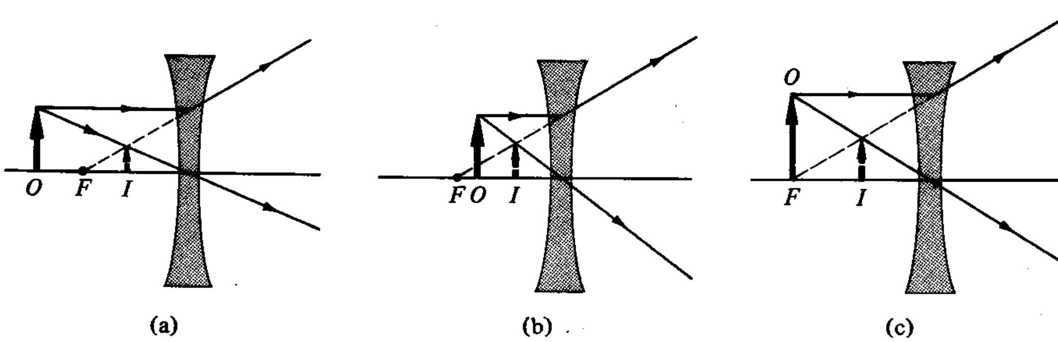


Figure 31-27

Example 31-11 Two thin lenses of focal lengths f_1 and f_2 are separated by a distance t; see Figure 31-28. Calculate the equivalent "focal length" of the combination.

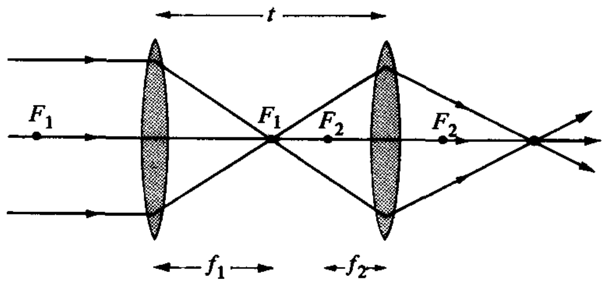


Figure 31-28

Solution To solve this problem, let us assume that parallel rays are incident on the lens of focal length f_1 and that the lenses are a distance t apart. By definition of focal length, these incident rays will form an image at a distance f_1 from the f_1-lens or, equivalently, at a distance $(t - f_1)$ from the second lens. On substituting this value $(t - f_1)$ for the object distance into (31-16) we obtain

$$\frac{1}{t - f_1} + \frac{1}{s'} = \frac{1}{f_2}$$

where s' is the image distance from the second lens. Calling this distance s' from the second lens, where the incident parallel rays come to a focus, the focal length f of the combination, we find that

$$\frac{1}{f} = \frac{1}{f_2} + \frac{1}{f_1 - t} \tag{31-22}$$

Thus by simply adjusting the distance t between the two lenses, it is possible to vary the focal length of this lens combination. The special case $t = 0$ corresponds to a "thin lens" of focal length f given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}$$

Note that if both lenses are convergent, then the focal length of the combination f is less than the focal lengths f_1 and f_2 of the constituent lenses.

31-9 Summary of important formulas

An object at a distance s from a concave spherical mirror of radius R will be imaged at a distance s' given by

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R} \tag{31-1}$$

and the lateral magnification m of the object will be

$$m = -\frac{s'}{s} \tag{31-9}$$

For a convex mirror the radius R in (31-1) must be replaced by its negative, $-R$. In these formulas, a positive value for s' means that the image is *real*

and *inverted*, and on the same side of the mirror as the object. Correspondingly, for $s' < 0$ the image is *virtual* and *erect*, and on the side of the mirror opposite to that of the object.

If an object is in a medium of refractive index n_1 and at a distance s_1 from a convex spherical interface with a medium of refractive index n_2, then an image will be formed at a distance s_2 from the interface, where

$$\frac{n_1}{s_1}+\frac{n_2}{s_2}=\frac{n_2-n_1}{R} \tag{31-10}$$

For a concave surface, the factor R in this formula must be replaced by $-R$. A positive value for s_2 means that a real image is formed in the n_2-medium, whereas a negative value for s_2 is associated with a virtual image in the n_1-medium.

If s is the distance of an object from a thin lens, then an image will be formed at a distance s', where

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f} \tag{31-16}$$

In this formula f is the focal length of the lens and is given by the lensmaker's equation

$$\frac{1}{f}=(n-1)\left[\frac{1}{R_1}+\frac{1}{R_2}\right] \tag{31-17}$$

where n is the refractive index of the lens, and R_1 and R_2 are positive provided that the centers of curvature of the two surfaces are oriented as in Figure 31-21. If either center of curvature is different than in this figure, the corresponding factor R_1 or R_2, or both, must be replaced by its negative. Positive values of s' correspond to a real image on the side of the lens opposite to the object, and negative values for s' imply that the image is virtual and on the same side of the lens as the object. The lateral magnification of a thin lens is given by the same formula as that for a spherical mirror in (31-9).

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** virtual image; **(b)** lateral magnification; **(c)** paraxial rays; **(d)** thin lens; **(e)** lensmaker's equation; and **(f)** converging lens.
2. A small object is pushed toward a plane mirror. Describe the motion of the image.
3. A point object lies somewhere between two plane mirrors which are at an angle of 60° to each other. Explain why you will see five images.
4. Consider two plane mirrors hinged along an edge and making an angle of 45°. Locate all seven images formed by a small object placed at the angle bisector of the mirrors. Characterize each image in terms of the number of reflections it undergoes.
5. What is the distinction between a real and a virtual image? Can a vir-

tual image be photographed? Can it be projected onto a screen?

6. A small object in front of a concave spherical mirror produces a real image. In view of the symmetry under exchange of s and s' in (31-1), how can you tell the difference between the object and the image only by viewing the light rays?
7. What are the approximations made in deriving the formula in (31-1) for a spherical mirror? Under what circumstances would this formula be *invalid*?
8. A point object is initially very far away from a concave spherical mirror of radius R. Describe the motion of the image as the object is brought along the axis to a final resting position at an object distance of $R/4$.
9. Repeat Question 8, but suppose this time that the mirror is convex.
10. Suppose that the object in Question 8 were a small pencil with its point directed perpendicular to the axis. Describe the variation in the size and orientation of the pencil image as the pencil is brought to within a distance of $R/4$ from the mirror.
11. Repeat Question 10, but this time assume that the mirror is convex.
12. Which would be a better choice for the side-mounted rear view mirror of an automobile: a concave mirror or a convex one? Explain your answer.
13. Imagine looking into a mirror which is very slightly curved and seeing a virtual, erect, and enlarged image of your face. Can you conclude from this observation that the mirror is concave? Explain.
14. In the problems it is shown that, for a concave *parabolic* mirror, if light rays, which are parallel to the axis of the mirror, are incident on it, then they come to a focus at a certain point on the axis. Explain in what way this differs from the corresponding case of parallel light rays incident on a concave *spherical* mirror.
15. What is the distinction between a *converging* and a *diverging* thin lens? Which one always produces a virtual image? Does one of these always yield a real image?
16. Where in front of a thin lens must an object be placed so that the image has a lateral magnification of -1? Must the lens be *converging*? Explain.
17. Why does the focal length of a thin lens increase if the lens is under water? Is this an important consideration for swimmers who wear contact lenses?
18. A thin, converging glass lens is placed into a transparent liquid whose refractive index exceeds that of glass. Explain why the lens becomes diverging under these circumstances! What would happen to the lens if it were double-concave?
19. Is it possible for the focal length of a thin lens to vary with the color of the light? Explain.
20. Explain in what sense the object and the image of a thin, converging lens are interchangeable. How could one tell the object and the image apart by viewing only?
21. Suppose you are suspended at a distance of 20 meters from the center of a large spherical room defined by a large spherical mirror having a radius of 50 meters. What would you see?
22. Light from the sun is focused by a thin, converging lens onto a piece of paper. Explain why the paper catches fire. Would this work with a thin, diverging lens? Explain.
23. A "zoom lens" is one whose focal length can be varied by means of a mechanical or electronic adjustment. Show how such a lens can be constructed by use of two or more thin lenses.

PROBLEMS

1. Consider an object 20 cm from a plane mirror. If you view the image from a distance of 50 cm directly behind the object (that is, 70 cm from the mirror), for what distance must you focus your eyes?
2. A small object is midway between two parallel, plane mirrors separated by a distance D.
   (a) Show that there are images in both mirrors at separation distances nD $(n = 1, 2, \ldots)$ from the object.
   (b) Draw a ray diagram to show how the image in one of the mirrors at a distance of $2D$ from the object is formed.
3. A concave spherical mirror used for shaving has a radius of 30 cm.
   (a) Where is the image, if the shaver's face is 10 cm from the mirror?
   (b) What is the magnification?
   (c) Is the image real or virtual? Erect or inverted?
4. A man looks into a convex spherical mirror of radius 30 cm. If his face is 10 cm from the vertex of the mirror:
   (a) Where does he see his image?
   (b) What is the magnification?
5. A reflecting telescope has a concave spherical mirror with a radius of 1.5 meters.
   (a) Where is the image of a rocket that is at a distance of 100 km from the telescope?
   (b) What is the magnification?
   (c) Assuming that the rocket is a sphere with a radius of 20 meters, what is the radius of its image?
6. A small object is 10 cm in length and is oriented perpendicular to the axis of a concave spherical mirror of radius 80 cm. If the object is 20 cm from the mirror:
   (a) Where is the image? Is it real or virtual?
   (b) What is the magnification?
   (c) What is the size of the image? Is it erect or inverted?
7. Repeat Problem 6, but this time suppose that the mirror is convex.
8. An object is at a distance s from the vertex of a concave spherical mirror of radius R.
   (a) If D represents the *sum* of the object and image distance (that is, $D = s + s'$), show that

$$D = \frac{s^2}{s - (R/2)}$$

   (b) For what value of s is D minimum? Consider only object distances greater than $R/2$.
9. Repeat both parts of Problem 8, but this time for a convex mirror. (*Note*: The formula for D in (a) must be modified for this case.)
10. Show that the lateral magnification m of an object that is at a distance s from a concave spherical mirror of radius R is given by

$$m = -\frac{R}{2s - R}$$

   What is the analogous formula for a convex mirror? What is the largest magnification achievable for this latter case?
11. An object of length l lies along the axis of a concave spherical mirror of radius R. If the nearer end of the object is at a distance s from the vertex of the mirror and if its length is very small compared to R, show that the length l' of the image is

$$l' = l\left(\frac{R}{2s - R}\right)$$

   Assume that $s > R/2$.

12. Consider in Figure 31-29 an object O at a distance s from a *convex* spherical mirror of radius R. Follow the procedure used to derive (31-1) for a concave mirror and thus establish the validity of (31-2).

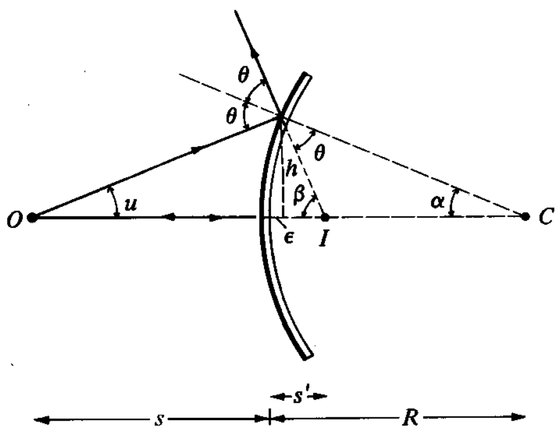


Figure 31-29

*13. Figure 31-30 shows a parabolic mirror whose equation in the given coordinate system is

$$y = \frac{a}{2}x^2$$

Consider an incoming ray traveling parallel to the y-axis along the line $x = x_0$ so that it strikes the mirror at the point $(x_0, ax_0^2/2)$.

(a) Show that the angle of incidence θ satisfies

$$\tan\theta = ax_0$$

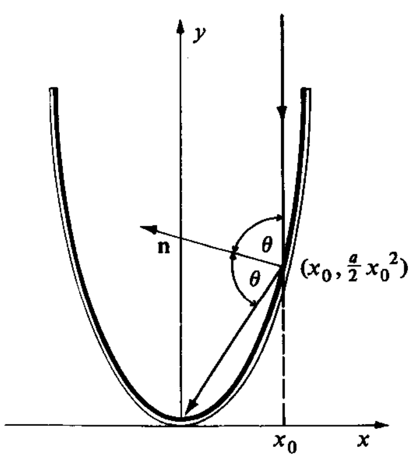


Figure 31-30

(b) Show that the equation for the reflected ray is

$$y = -(x - x_0)\cot 2\theta + \frac{a}{2}x_0^2$$

(c) Show that this reflected ray intersects the y-axis at a distance $1/2a$ above the origin. Explain why this result implies that the point with coordinates $(0, 1/2a)$ is the focal point of this parabolic mirror.

14. (a) An object of height 10 cm is perpendicular to the axis of a concave spherical mirror of radius 50 cm and is at a distance 60 cm from its vertex. Make a drawing of the situation to some scale, and determine the location and size of the image graphically. (*Hint*: Any ray through the center of curvature of the mirror is reflected back on itself, whereas any ray that goes through the focal point is reflected parallel to the axis.)

(b) Repeat (a), but this time assume that the object is 40 cm from the mirror.

15. Repeat (a) and (b) of Problem 14, but suppose this time that the mirror is convex.

16. It is desired to project the image of an object 3 cm high onto a screen 40 meters from a spherical concave mirror so that the size of the image is 80 cm.

(a) What lateral magnification is required?

(b) What must be the object distance s?

(c) What must be the radius of curvature of the mirror?

17. Show that for a plane mirror the *longitudinal magnification* (ds'/ds) is the same for all object distances, and determine its numerical value.

18. An object 15 cm high is in air and at a distance of 30 cm from a convex

spherical glass surface ($n = 1.5$) of radius 60 cm. Find the position of the image and state whether it is real or virtual and whether it is erect or inverted.

19. A goldfish is 5 cm long and is at the center of a spherical bowl of radius 40 cm, which is full of water ($n = 1.33$). Neglect refraction due to the bowl itself.
   (a) Where does the fish appear to be to an external observer?
   (b) Does the fish appear to be erect or inverted? (Assume that the fish is perpendicular to the line from the observer's eye to the center.)

20. A point source O is at the bottom of a tank of water of depth h and refractive index 4/3. Suspended at a distance d above the tank is a plane mirror parallel to the surface of the water; see Figure 31-31. Where would an observer who looks into the mirror see the image of the object?

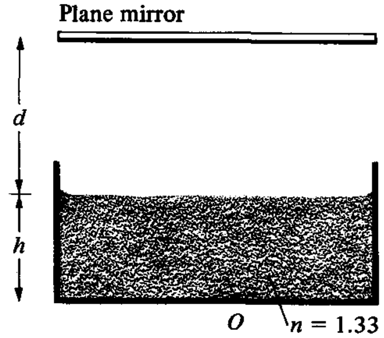


Figure 31-31

***21.** Repeat Problem 20, but suppose that this time the plane mirror is replaced by a concave spherical one of radius R and that d represents the distance from the vertex of the mirror to the surface of the water. Assume, $d + h = R/2$.

22. A bundle of parallel light rays traveling in glass ($n = 1.5$) are incident at its spherical interface with water ($n = 1.33$). Describe the nature and location of the image if:
   (a) The surface is concave of radius R.
   (b) The surface is convex of radius R.

23. Repeat both parts of Problem 22, but suppose that this time the original light rays are traveling in water and are refracted by the spherical glass surface.

24. Suppose that in Figure 31-17 the object O has a height y_1 and is perpendicular to the axis. If y_2 is the corresponding size of the image, show that the lateral magnification defined as $m = -y_2/y_1$ is

$$m = -\frac{n_1 s_2}{n_2 s_1}$$

25. A glass rod of refractive index 1.5 has a length of 60 cm and has at its ends hemispherical surfaces, each of radius 10 cm. An object of height 2 cm is placed at a distance of 30 cm from one end, as shown in Figure 31-32.

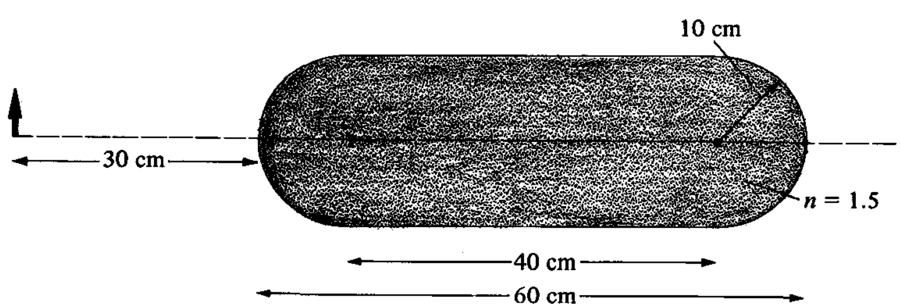


Figure 31-32

(a) Where is the image I_1 located after refraction by the left surface?
(b) What is the nature and size of this image I_1? (*Hint*: Use the result of Problem 24.)
(c) Find the location of the final image I_2, which is formed by the refraction of the light from I_1, considered as an object for the second surface.
(d) What is the nature and magnification of I_2 relative to the original object?

26. Consider the rod in Figure 31-32, but suppose that this time light rays parallel to its axis are incident on its left surface.
(a) Where does the light come to a focus after the first refraction?
(b) What is the location of the final image after two refractions?

***27.** Parallel beams of light are incident on a glass sphere of radius R and index 1.5. If, as shown in Figure 31-33, a plane mirror is at a distance (3/2)R from the sphere, calculate the location of the final image. (*Hint*: See Example 31-10.)

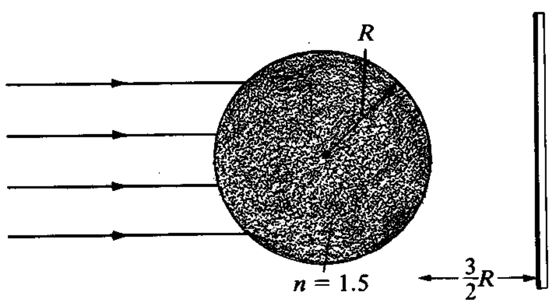


Figure 31-33

***28.** Consider again the situation in Figure 31-33, but suppose that this time an object is located at a distance 3R from the left surface of the sphere.
(a) Find the location of the image after the first refraction through the surface of the sphere.
(b) Find the final position of the image.

29. A plano-convex thin lens is made of glass ($n = 1.5$) and the curved surface has a radius of 10 cm.
(a) What is the focal length of the lens?
(b) Repeat (a), but suppose that this time the lens is plano-concave with the same radius of curvature.
(c) Is the lens in (a) diverging or converging?

30. A double convex thin lens made of glass ($n = 1.5$) has a focal length of 30 cm. If the radius of one of the surfaces is 30 cm, what is the radius of the other surface?

***31.** Show that if f_0 is the focal length of a thin lens in air and f' is the corresponding focal length when the lens is in a medium of refractive index n', then

$$f' = f_0 \frac{n'(n-1)}{(n-n')}$$

where n is the refractive index of the material of the lens. (*Hint*: Modify the arguments used to derive the lensmaker's equation.)

32. A thin, converging lens has a focal length of 20 cm. Find the location and magnification associated with an image whose object distance is: **(a)** 30 cm; **(b)** 20 cm; **(c)** 10 cm. In which of these cases is the image real?

33. An object is at a distance of 12 cm from a thin, converging lens of focal length 10 cm. **(a)** How far from the lens must a screen be placed so that the image is focused on it? **(b)** How large is the image if the object is perpendicular to the axis and has a size of 4 cm?

34. An object is at a distance d from a screen. Show that if a thin converging lens of focal length f ($< d/4$) is placed at either of the distances

$$\frac{d}{2} \pm \frac{1}{2}[d(d-4f)]^{1/2}$$

from the object, then the image will come to a focus on the screen. What is the ratio of the lateral magnification in the two cases?

35. Show that if an object is at a distance $(x+f)$ from a thin lens of focal length f, then its image is at a distance $(x'+f)$, where

$$xx' = f^2$$

(*Note*: This form of the thin lens equation is known as the *Newtonian form* and is to be contrasted with the *Gaussian form* in (31-16).)

36. An object is at a distance s from a thin lens of focal length f.

(a) Show that the distance d between the object and the image is

$$d = \frac{s^2}{s-f}$$

(b) Show that for a converging lens the distance between the object and any real images formed is always greater than $4f$.

(c) At what object distance is the minimum in (b) achieved?

37. When an object having a height of 3 cm is placed 8 cm in front of a thin, converging lens, an image is formed on a screen 12 cm behind the lens. Suppose that the lens is moved 2 cm closer to the object. **(a)** Which way must we move the screen so that the image is still in focus? **(b)** What is the final size of the image?

38. An object 2 cm high is at a distance of 10 cm from a converging lens of focal length 4 cm. Make a drawing to scale and, by drawing appropriate rays from the top of the object, determine the size and location of the image. Compare your graphical values with the calculated ones.

39. Repeat Problem 38 but suppose that this time the lens is diverging and has a focal length of -4 cm. Assume that the object is 4 cm from the lens.

40. Two converging lenses each of focal length 10 cm are separated by a distance of 50 cm. An object 5 cm in height is placed to the left of the lenses and 15 cm from one of the lenses and 65 cm from the other.

(a) Where is the image I_1 due to the first lens?

(b) Where is the final image I_2 located?

(c) What are the sizes of I_1 and I_2? Is the final image erect or inverted?

41. Consider the lens system of Problem 40, but suppose that this time the more remote lens is moved so that it is only 35 cm from the other. What is the nature and position of the final image?

42. In Figure 31-28 suppose that $f_1 = +10$ cm, $f_2 = +10$ cm, and $t = 50$ cm. How far from the second lens does the light come to a focus?

43. Repeat Problem 42, but suppose that this time $f_1 = -10$ cm, $f_2 = +10$ cm, and $t = 50$ cm.

44. Figure 31-34 shows the basic ele-

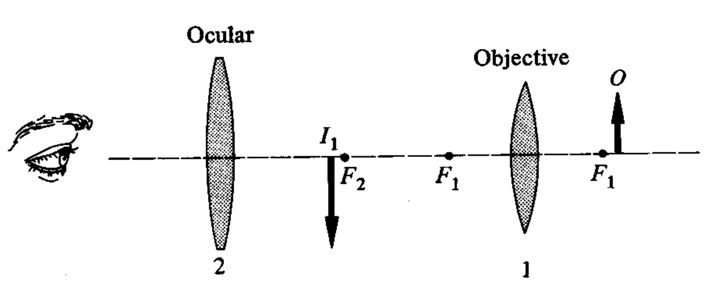


Figure 31-34

ments of the compound microscope. The object O is placed just outside the focal point of the objective lens and the associated image falls just inside the focal length of the ocular lens. Explain why the observer will see an enlarged image of the object. Will the observed image be real or virtual? Erect or inverted?

45. The basic elements of a telescope are two converging lenses: an *objective* lens (1) of focal length f_1 and an *ocular* lens (2) of focal length f_2. Unlike the microscope in Figure 31-34, for a telescope the lenses are separated by the distance $(f_1 + f_2)$. Explain, by constructing a ray diagram for an object at infinity, how the telescope works. Show also that the magnification M of a telescope is

$$M = -\frac{f_1}{f_2}$$

and explain the significance of the minus sign.

46. An object is at a distance $(a + b)$ from a plane mirror. Suppose that a thin, converging lens of focal length f $(f < a;\ f < b)$ is placed between the object and the mirror and at a distance a from the former. Find the nature and the location of the final image assuming that $b > af/(a - f)$. Justify your answer by constructing an appropriate ray diagram.

***47.** A point object O is in front of a concave spherical mirror of radius R, and forms a real image I. Prove that the time it takes for any two paraxial rays to go from O to I is the same. What is the relation between this result and Fermat's principle as described in Problems 7 and 8 of Chapter 30? (*Hint*: Make use of the geometry of Figure 31-11 and show that for paraxial rays the time it takes the ray OA to travel the path OAI is independent of the angle u.)

***48.** Suppose, in Figure 31-17, that an object O is near an interface between two media of refractive indices n_1 and n_2, and a *real* image I is formed. Prove that *all* paraxial rays take the same time to go from O to I regardless of the values of R, n_1, and n_2. How is this result related to Fermat's principle?

49. By making use of the results of Problem 48 or of Fermat's principle, show that if a real image I is formed of an object O by a thin, positive lens, then the time taken in going from O to I is the same for *all* paraxial rays.

50. Suppose, in Figure 31-35, that parallel rays are incident at an angle u with respect to the axis of a thin, positive lens of focal length f. Show that the rays will converge to a point image I at a distance f behind the lens. The locus of the images formed in this way by various incident angles u defines the *focal plane* of the lens. (*Hint*: Consider the incident rays that go through the focal point and through the center of the lens.)

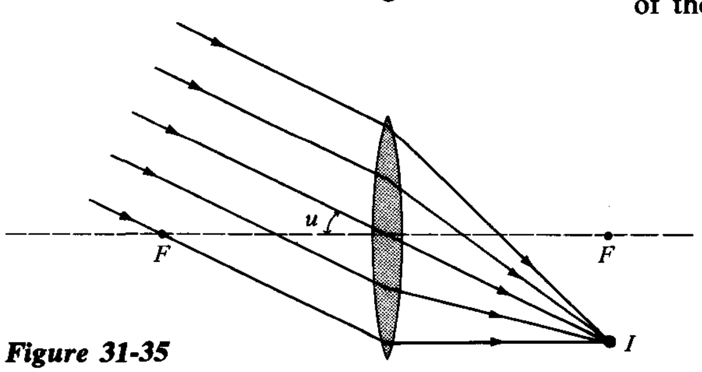


Figure 31-35

32 Interference and diffraction

Where there is a great deal of light, the shadows are deeper.
GOETHE

32-1 Introduction

The fact that light is a wave motion did not play an important role in our formulation of the laws of *geometrical optics* in Chapter 30. The reason for this is that these laws are applicable only to physical situations involving the interaction of light with objects whose dimensions are large compared to a wavelength. Under these conditions the fact that light is a wave motion plays only a very minor role.

By contrast, in this chapter we shall consider the entirely new set of phenomena that arises when light interacts with objects whose linear dimensions are comparable to a wavelength. This is the field of *physical optics*. The new physical effects here are: (1) the *diffraction* of light as manifested by the bending of light around the edges of sharp obstructions; and (2) the *interference* of two light beams for which the observed intensity is *not* the sum of the intensities of the separate beams. As we shall see, diffraction and interference are two very closely related phenomena. Although Christian Huygens suggested the possibility of the diffraction and the interference of visible light as early as 1690, it was not until 1803 that

Thomas Young (1773–1829) succeeded in establishing the validity of this hypothesis in the laboratory. His key experiment was simplicity itself. In his own words:

> I made a small hole in a window shutter and covered it with a piece of thick paper, which I perforated with a fine needle . . . . I placed a small looking-glass without the window shutter in such a position as to reflect the sun's light, in a direction nearly horizontal, upon the opposite wall . . . . I brought into the sunbeam a slip of card about one thirtieth of an inch in breadth and observed its shadow, either on the wall or on other cards held at different distances. Besides the fringes of colors on each side of the shadow, the shadow itself was divided by similar parallel fringes of smaller dimensions, differing in number according to the distance at which the shadow was observed but leaving the middle of the shadow always white. Now these fringes were the joint effects of the portions of light passing on each side of the slip of card, and inflected, or rather diffracted into the shadow . . . .

32-2 Qualitative aspects of diffraction

The term diffraction refers to the experimental fact that light rays when passing around the sharp edge of an object will spread out, to some extent, into the region of its shadow. Because the wavelength of visible light is of the order of 5×10^{-5} cm and is thus small compared to the dimensions of ordinary bodies this effect is difficult to observe except under controlled laboratory conditions.

Consider, in Figure 32-1, a parallel beam of light rays incident at right angles onto a barrier AB, out of which has been cut an aperture H. Suppose first that the linear dimensions, a, of the aperture are of the order of a centimeter and thus very large compared to a wavelength. According to the ideas of geometrical optics, the incident light rays that strike the barrier are either absorbed or scattered by it, whereas those that are incident at the opening will go through undeviated. Thus we might expect that a bright spot will appear at EF, and that the remainder of the screen—that is, CE and DF—will remain dark. Indeed, provided that $a \gg \lambda$, these expectations are substantially in accord with observations. However, even for this case in which the linear dimensions of the aperture are very large compared to a wavelength, a careful examination of the edges of the bright spot at E and F shows that the line of demarcation between the bright spot and the shadow on the screen is not very sharp. But rather, very close scrutiny shows that near E and F there are some dark bands inside the bright region and some bright bands in the region of the shadow.

Imagine now repeating the above experiment, but with successively smaller apertures. The dividing line between the bright and the dark region on the screen will become less and less sharp until a situation such as that in Figure 32-2 is reached. Here the incoming light is incident on a tiny opening

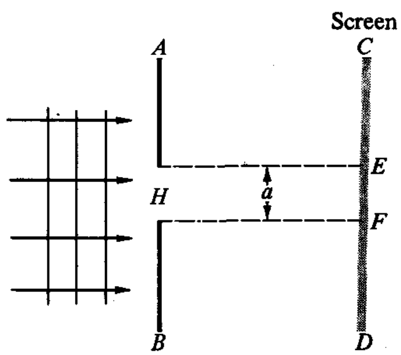


Figure 32-1

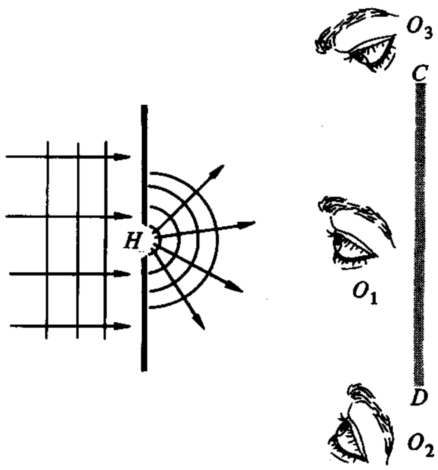


Figure 32-2

whose linear dimensions are small compared to the wavelength of visible light. The pattern of illumination on the screen under this circumstance is completely at variance with the ideas of geometrical optics. As shown in the figure, the observed pattern is best described as one that results if spherical waves—that is, waves whose surfaces of constant phase, or *wavefronts*, are spherical surfaces—emanate radially outward from the aperture. If geometrical optics were applicable to this situation, then only the observer O_1 who is lined up with the aperature H would see any light. In fact, we find that the light emanating from H is also seen by various other observers, such as O_2 and O_3, thus confirming that the light from H spreads out after going through the opening in the barrier. This phenomenon of the spreading out of a light beam after passing through an aperture of dimension of the order of 10^{-6} meter and smaller is an illustration of the *diffraction* of light; it is a direct and unambiguous manifestation of the fact that light is a wave motion.

Figure 32-3 illustrates the kinds of patterns observed for the analogous case of water waves in a ripple tank passing through an opening. Note that if the size of the opening is very large compared to a wavelength (the distance between neighboring crests or troughs), there is relatively little diffraction. On the other hand, for a narrow opening, the emerging rays are the radii of circles centered at the opening.

A qualitative way of thinking of the phenomenon of diffraction can be obtained by use of a rule known in honor of its discoverer as the *Huygens construction.* According to this rule, the wavefront of a wave at any instant $(t + \Delta t)$ may be obtained from that at a previous instant t by thinking of each point of the earlier wavefront as being a point source of outgoing spherical waves. The envelope of these spherical waves is then the wavefront at the later time $(t + \Delta t)$. By way of example, Figure 32-4a shows a portion AB of an "infinite" wavefront at some instant t. The associated wavefront $A'B'$ at the later instant $(t + \Delta t)$ is the envelope, or the common tangent plane in this case,

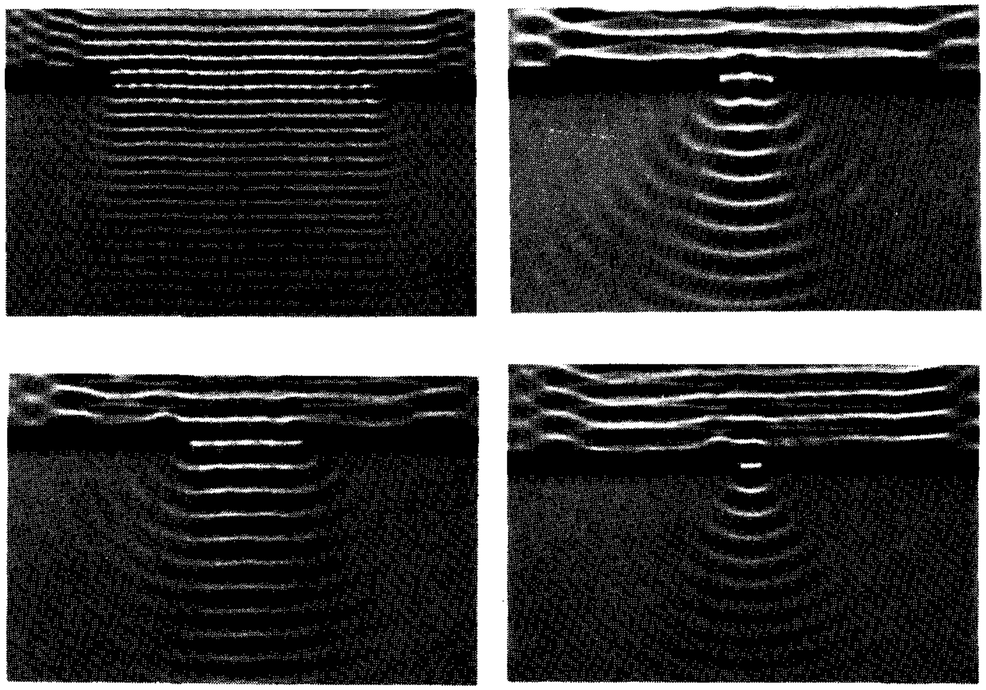
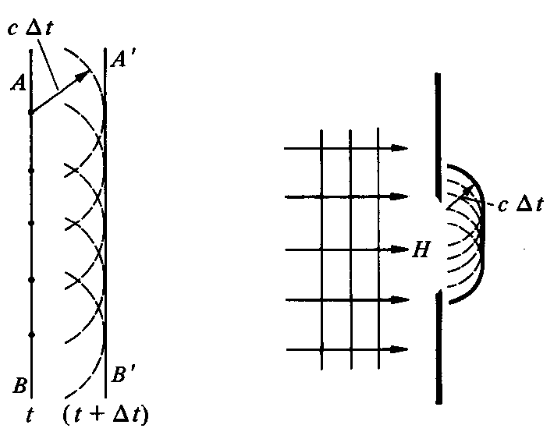


Figure 32-3 Water waves being diffracted through slits of varying sizes in a ripple tank. Note how the diffraction pattern widens as the size of the slit is decreased down to a wavelength. (Courtesy of Project Physics, Holt, Rinehart and Winston, New York.)

of the spherical wavefronts, each of radius $c\ \Delta t$ and emanating from the various points of the original wavefront AB. It is evident that, consistent with our expectations, the new wavefront $A'B'$, as given in this way by the Huygens construction, is parallel to AB and at a distance $c\ \Delta t$ from it. Figure 32-4b illustrates the corresponding situation of a wave of infinite extent approaching an aperture H. Applying the Huygens construction to the finite wavefront which is at the aperture at time t, we may construct the new wavefront at the later instant $(t + \Delta t)$. Note that only near the center of the aperture is the transmitted wavefront parallel to the original one. At the edges the transmitted light has been diffracted.



(a) (b)

Figure 32-4

Despite its historical interest and simplicity, the Huygens construction will not be used in the following quantitative description of diffraction. Instead, we shall take the point of view that light is an electromagnetic wave and then establish that diffraction and interference are properties of all such waves.

32-3 A point source

By way of introduction to a quantitative study of interference, in this section we review briefly some of the important parameters associated with electromagnetic waves emitted by a single source.

Consider, in Figure 32-5, a monochromatic point source S, which radiates electromagnetic waves of wavelength λ. As shown, very close to the source the wavefronts of the outgoing radiation are spherical, but to an observer O who is very far from S these wavefronts will appear to be plane waves. If E_0 is the amplitude of the observed waves, then, according to (29-20) through (29-23), the electric field E in the radiation seen by the observer O is

$$E = E_0 \sin\left[\frac{2\pi}{\lambda}(z - ct) - \alpha\right] \tag{32-1}$$

where $c = 3.0 \times 10^8$ m/s is the speed of light and α is a constant phase factor. In writing down this formula, we have assumed that the source is at the origin of the coordinate system and that the observer is at a remote distance z along the z-axis. Since the **E** and **B** vectors in an electromagnetic wave are always perpendicular to each other and to the propagation direction of the wave, a knowledge of only the magnitude of the electric field will be needed in the following. The magnetic induction **B** associated with the wave in (32-1) is given by (29-20), but with the argument of the sine function the same as that in (32-1).

The argument of the sine function in (32-1), that is, the factor $[2\pi(z - ct)/\lambda - \alpha]$, is known as the *phase* of the wave. All points along a given wavefront have the same phase. In particular, for a plane wave all points of the wave which lie in a fixed plane at right angles to the propagation direction of the wave are characterized by the same phase.

The magnitude N of the Poynting vector—that is, the flow of energy per

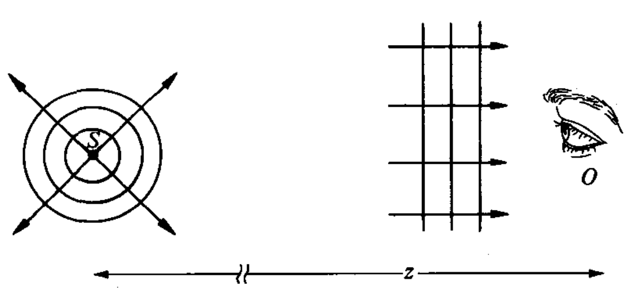


Figure 32-5

unit area per unit time received by the observer—is obtained by substituting (32-1) and the associated formula for **B** into (29-17). The result is

$$N = \sqrt{\frac{\epsilon_0}{\mu_0}}\, E_0^2 \sin^2\left[\frac{2\pi}{\lambda}(z - ct) - \alpha\right] \tag{32-2}$$

As in Chapter 29, only the average value I_0 of this quantity over a period λ/c is of physical interest. Proceeding, therefore, as in the analogous derivation of (29-25), we find for the average intensity

$$I_0 = \frac{1}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\, E_0^2 \tag{32-3}$$

Note that I_0 is independent of both the phase α and the wavelength λ of the electromagnetic wave.

32-4 Interference

Consider, in Figure 32-6, two identical point sources S_1 and S_2, each emitting electromagnetic waves of the same amplitude and of the wavelength λ. As for the single-source case considered above, suppose the observer O is very far away from the sources, and to simplify matters let us assume that S_1, S_2, and O are along a straight line. We define the *path difference* between two sources and the given observer to be the difference in the optical paths between the sources and the observer. For the situation in Figure 32-6 this means that since S_1, S_2, and O are collinear, the distance Δ between the sources is itself the path difference. In other words, since the distance between S_1 and O is z while that between S_2 and O is $(z-\Delta)$, the path difference is $z-(z-\Delta)=\Delta$. In the following we shall be concerned mainly with physical situations for which the path difference Δ is of the order of the wavelength of light. By contrast, the distance z will always be assumed to be so large that the wavefronts seen by the observer are planes, so that in general $|z| \gg \Delta$.

Let us now calculate the average intensity at the position of the observer. According to the superposition principle, the electric field E in the wave seen by the observer is the sum of the fields produced by each of the sources separately. Hence, taking note of (32-1) and neglecting effects associated

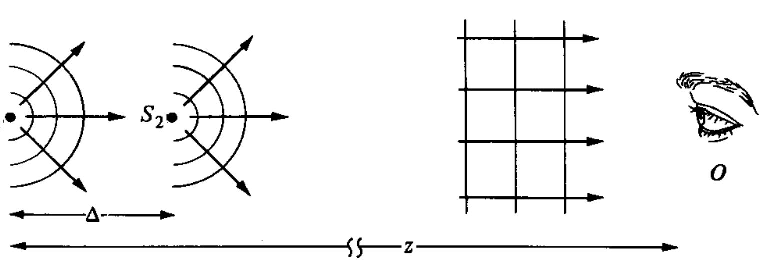


Figure 32-6

with the vector nature of the electric field, we find that

$$E = E_0 \sin\left[\frac{2\pi}{\lambda}(z - ct) - \alpha_1\right] + E_0 \sin\left[\frac{2\pi}{\lambda}(z - \Delta - ct) - \alpha_2\right] \quad \textbf{(32-4)}$$

with E_0 the observed electric field amplitudes, and α_1 and α_2 certain phase factors. Note that even though $\Delta \ll |z|$, the path difference Δ cannot be neglected compared to z in this formula.

To explore the possibility of interference, suppose first that the two sources are independent of each other. In this case there is no definite phase relation between them or, in other words, the phases α_1 and α_2 in (32-4) are *random*. We say in this case that the sources are *incoherent*. For incoherent sources the observed intensity I is the sum of the intensities due to each source. Hence, since S_1 and S_2 have the same amplitude E_0, it follows that if they are incoherent, then

$$I = 2I_0 \qquad \text{(incoherent sources)} \quad \textbf{(32-5)}$$

with I_0 given by (32-3). Incoherent sources, such as the light coming from different points on this page, *do not interfere* with each other.

By contrast, if for some reason the phase difference $(\alpha_1 - \alpha_2)$ is constant over long periods of time, then we say that the two sources are *coherent*. Assuming for simplicity that $\alpha_1 = \alpha_2$, it will be established below that the average intensity I produced by two coherent sources is

$$I = 4I_0 \cos^2\frac{\delta}{2} \qquad \text{(coherent sources)} \quad \textbf{(32-6)}$$

with I_0 defined by (32-3). The new quantity δ, which will play an important role in the following, is defined by

$$\delta = \frac{2\pi}{\lambda}\Delta \quad \textbf{(32-7)}$$

and is known as the *phase difference* between the coherent sources.

According to (32-6), the observed intensity I for coherent sources can vary from a maximum of $4I_0$ to a minimum of zero, depending on the path difference Δ between the sources. If Δ is an integral number of wavelengths, for example, λ or 2λ, then the phase difference δ is an integral multiple of 2π and, according to (32-6), I achieves its maximum value of $4I_0$. We say in this case that *constructive interference* has taken place. On the other hand, if Δ is an odd integral multiple of $\lambda/2$, such as $\lambda/2$, $3\lambda/2, \ldots$, then δ is an odd multiple of π and I achieves its minimum value of zero. In this case we say that *destructive interference* has taken place.

Figure 32-7 shows a graph of (32-6) as a function of the phase difference δ. Note that, as shown above, at $\delta = \pi, 3\pi, 5\pi, \ldots$, the interference is destructive, whereas at $\delta = 0, 2\pi, 4\pi, \ldots$ it is constructive. The horizontal dotted line at $I = 2I_0$ represents the corresponding plot for the case of incoherent sources.

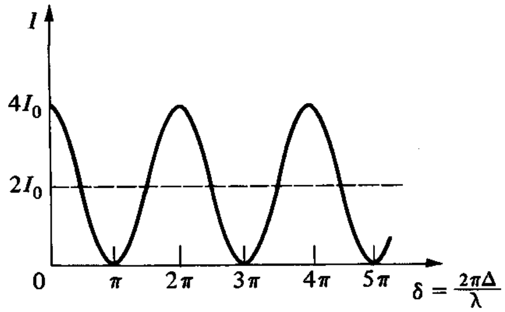


Figure 32-7

It is important and interesting to note that even for the case of complete destructive interference there is no contradiction with the principle of energy conservation. For even though the path difference Δ between the sources relative to the observer O in Figure 32-6 is an odd integral multiple of half of a wavelength, there are other observation points relative to which Δ is an integral multiple of a wavelength. For example, to an observer who views the sources S_1 and S_2 along the perpendicular bisector of the line joining them, the path difference of the sources relative to him vanishes. Hence, according to (32-6) and (32-7), he would see the maximum intensity $I = 4I_0$. As we might expect, the total intensity radiated by the two sources in *all* directions is precisely the same as that which would be radiated if they were incoherent.

Figure 32-8 shows the analogous interference pattern when water waves pass through neighboring holes in a ripple tank. Note that there are directions along which the amplitudes of the wave are a maximum and others along which they are minimum. The path difference associated with the former is invariably an integral multiple of λ and with the latter an odd multiple of $\lambda/2$.

To derive (32-6) let us make use of the trigonometric identity

$$\sin A + \sin B = 2\cos\left(\frac{A-B}{2}\right)\sin\left(\frac{A+B}{2}\right)$$

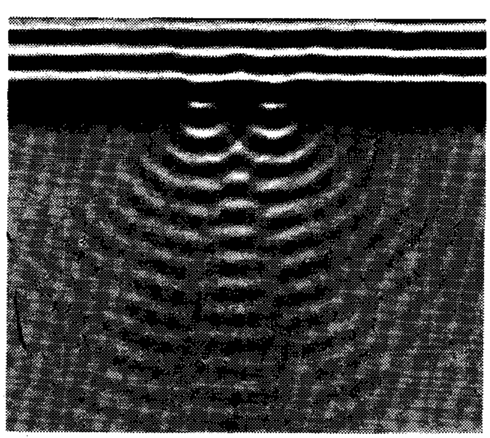


***Figure 32-8** An interference pattern produced in a ripple tank by the diffraction of water waves from two nearby slits. (Courtesy Project Physics, Holt, Rinehart and Winston, New York.)*

to reexpress the formula for E in (32-4) in the form

$$E = 2E_0 \cos\frac{\delta}{2} \sin\left[\frac{2\pi}{\lambda}(z - ct) - \frac{\delta}{2} - \left(\frac{\alpha_1 + \alpha_2}{2}\right)\right]$$

Here δ is the phase difference defined in (32-7), and use has been made of the fact that the sources are coherent and that $\alpha_1 = \alpha_2$. Comparing this form for E with that in (32-1), we see that it corresponds to the radiation emitted by a single source with amplitude $2E_0 \cos(\delta/2)$. Hence substitution into (32-3) yields the desired formula in (32-6).

The formal derivation of (32-5) for the radiation emitted by two incoherent sources proceeds in a similar way. It is left as an exercise to confirm that (32-5) results if an averaging process over the random phases α_1 and α_2 is carried out.

Example 32-1 Two point sources are a distance 1.0×10^{-6} meter apart and are emitting light of wavelength 5500 Å. Assuming that the observed intensity when either source is acting alone is 2.0 W/m^2, and that the observer is along the line joining the sources, calculate:

(a) The observed intensity if the sources are incoherent.
(b) The observed intensity if they are coherent.
(c) The maximum intensity possible for these coherent sources and the smallest additional distance they must be separated to achieve this maximum.

Solution

(a) The substitution of the given value $I_0 = 2.0\ \text{W/m}^2$ into (32-5) yields for the incoherent case

$$I = 2I_0 = 4.0\ \text{W/m}^2$$

(b) For the coherent case, we must use (32-6). According to (32-7) the phase difference δ is

$$\delta = \frac{2\pi}{\lambda}\Delta = \frac{2\pi \times 1.0 \times 10^{-6}\ \text{m}}{5.5 \times 10^{-7}\ \text{m}} = 3.6\pi\ \text{rad}$$

and hence

$$I = 4I_0 \cos^2\frac{\delta}{2} = (4 \times 2.0\ \text{W/m}^2) \times \cos^2 1.8\pi = 5.2\ \text{W/m}^2$$

(c) The intensity will be maximum at the value $4I_0 = 8.0\ \text{W/m}^2$ when the conditions for constructive interference are satisfied. This will take place when the path difference Δ is an integral number of wavelengths. Initially the distance between the sources is

$$1.0 \times 10^{-6}\ \text{m} = 1.8 \times 5.5 \times 10^{-7}\ \text{m} = 1.8\lambda$$

Hence they must be separated an additional distance of

$$(2 - 1.8)\lambda = 0.2\lambda = 0.2 \times 5.5 \times 10^{-7}\ \text{m} = 1.1 \times 10^{-7}\ \text{m}$$

to achieve this maximum.

Example 32-2 Suppose that the two sources in Figure 32-6 are emitting white light—that is, light of all wavelengths. If the separation distance between the sources

is the minimum one for which blue light of wavelength 4500 Å suffers destructive interference, what wavelength will appear with maximum intensity?

Solution Reference to the graph in Figure 32-7 shows that for the blue light the phase difference δ must be π. Thus, according to (32-7) the value of Δ must be

$$\Delta = \frac{\lambda\delta}{2\pi} = \frac{4500\pi \text{ Å}}{2\pi} = 2250 \text{ Å}$$

And since a maximum in intensity I will occur when Δ is an integral number of wavelengths, it follows that the wavelength of the light associated with this maximum is

$$\lambda = 2250 \text{ Å}$$

In other words, radiation at this wavelength in the ultraviolet region will be at a maximum intensity for the given separation distance between the sources.

Example 32-3 Generalize (32-6) to the case of N identical coherent sources S_1, $S_2, \ldots, S_N$ located at the points with z-coordinates of $0, \Delta, 2\Delta, \ldots, (N-1)\Delta$, respectively. Assume that the observer is at a remote point along the z-axis and that $|z| \gg (N-1)\Delta$, so that he sees plane waves.

Solution Since the sources are coherent, the generalization of (32-4) for N sources is

$$E = E_0 \sum_{i=0}^{N-1} \sin\left[\frac{2\pi}{\lambda}(z - ct - i\Delta) + \alpha\right]$$

provided that the phases α are all the same. Making use of the following trigonometric identity (derived in Problem 13)

$$\sum_{i=0}^{N-1} \sin(x + i\theta) = \frac{\sin(N\theta/2)}{\sin(\theta/2)} \sin\left[x + \frac{(N-1)\theta}{2}\right] \tag{32-8}$$

this formula for E may be reexpressed in the form

$$E = E_0 \frac{\sin(N\pi\Delta/\lambda)}{\sin(\pi\Delta/\lambda)} \sin\left\{\frac{2\pi}{\lambda}\left[z - ct - \left(\frac{N-1}{2}\right)\Delta\right] + \alpha\right\}$$

Comparison with (32-1) and (32-3) thus yields for the observed intensity the result

$$I = I_0 \frac{\sin^2[N\pi\Delta/\lambda]}{\sin^2[\pi\Delta/\lambda]} \tag{32-9}$$

with I_0 the intensity which would be observed if only one source were radiating. It is left as an exercise to confirm that for $N = 2$ this reduces to (32-6), as it must.

32-5 Young's two-slit experiment

In order to verify experimentally the existence of interference phenomena as predicted by (32-6) and (32-7), it is necessary to have available coherent sources. In the next two sections we shall describe several methods that

have been used in the past to produce such sources. These methods, although of considerable historical and conceptual interest, are no longer as widely used as they used to be. The discovery of the *laser,* which is a device capable of producing a very intense and highly coherent light beam, has made them obsolete.

The classic method used to demonstrate interference was devised by Thomas Young. Figure 32-9 shows the essential features of his experimental setup. Light from a line source S (perpendicular to the plane of the figure) is incident through a converging lens L onto a barrier AB out of which have been cut two slits S_1 and S_2. Assuming that S is in the focal plane of the lens, the cylindrical wavefronts emanating from S will approach the barrier as plane waves with wavefronts parallel to AB. (Equivalently, the source S and the lens L may be replaced by a laser beam.) Since, by definition, all points of a wavefront have the same phase, it follows that the waves that emerge from the slits will be coherent and, if allowed to fall on a screen, will produce an interference pattern. As is implied in the figure, the screen is assumed to be far enough away from the slits so that the waves incident on the screen are plane waves. For only then can the analysis of Section 32-4 be used. Alternatively, if the waves coming from S_1 and S_2 are first refracted through a thin, converging lens, then the screen can be moved up to coincide with the focal plane of this lens. (See Problem 50 in Chapter 31 and the corresponding analysis of the diffraction grating in Section 32-7.)

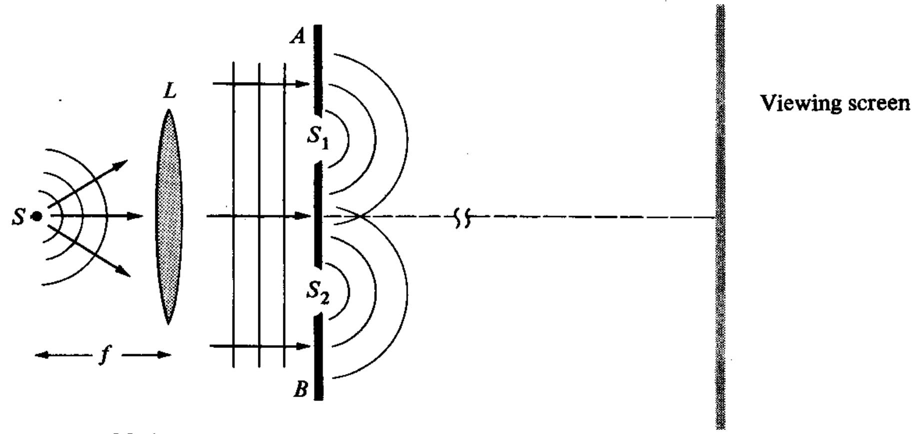


Figure 32-9

To analyze the interference pattern on the screen, let d represent the distance between the slits and D ($\gg d$) the distance from the slits to the screen. Draw a reference line perpendicular to and bisecting the line joining the slits and let it intersect the screen at P_0, as in Figure 32-10. Any point P on the screen can then be described in terms of its distance x from P_0 or, equivalently, in terms of the angle θ between the lines going from the midpoint of the slits to P_0 and to P.

Consider now the two rays S_1P and S_2P, which go from the slits to the point P. At S_1 and S_2 they are in phase, so their path difference, on arriving at

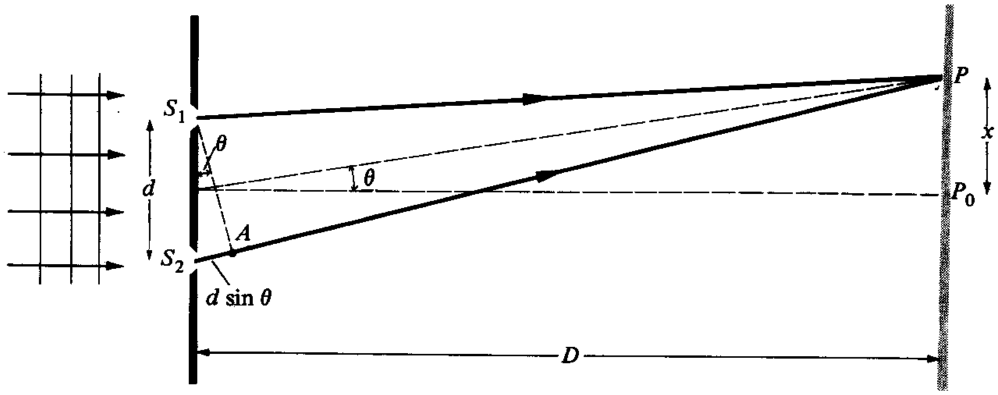


Figure 32-10

P, is the difference $(S_1P - S_2P)$. Since D is presumed to be very large and d very small, it follows that the rays S_1P and S_2P are essentially parallel. Hence the path difference Δ is effectively the distance S_2A, with A determined by drawing a perpendicular from S_1 to the ray S_2P. Reference to the figure shows that the angle between S_1S_2 and S_1A is also θ. Therefore the path difference between the two rays S_1P and S_2P is $d \sin \theta$ and substitution into (32-7) yields

$$\delta = \frac{2\pi d}{\lambda} \sin \theta \qquad \textbf{(32-10)}$$

The corresponding formula for the intensity $I(\theta)$ observed on the screen at the angle θ can be found by substitution into (32-6) with the result

$$I(\theta) = 4I_0 \cos^2 \left[\frac{\pi d \sin \theta}{\lambda}\right] \qquad \textbf{(32-11)}$$

where $4I_0$ is a constant which represents the maximum intensity on the screen.

Figure 32-11 shows a plot of the intensity $I(\theta)$ on the screen as a function of the quantity $\delta/2 = (\pi d \sin \theta)/\lambda$. Note that at the values for θ defined by

$$\sin \theta_m = \frac{m\lambda}{d} \qquad (m = 0, \pm 1, \pm 2, \ldots) \qquad \text{(maxima)} \qquad \textbf{(32-12)}$$

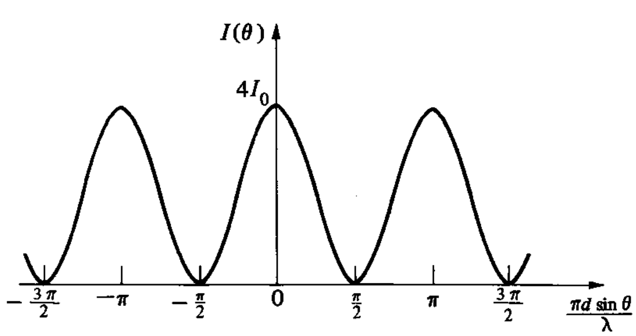


Figure 32-11

the intensity in the pattern is a maximum. The bright line corresponding to $m = 0$, which occurs at the center of the screen at P_0, is known as the *zeroth-order* fringe, and the neighboring lines at $\pm\theta_1$ corresponding to $m = \pm 1$ are called the first-order fringes, and so forth. The intensity maximum corresponding to θ_m appears at that position on the screen for which the path difference from S_1 and S_2 is $m\lambda$. Between these intensity maxima there are dark bands associated with positions on the screen for which the path difference is an odd integral multiple of $\lambda/2$. These minima occur for values of the angle θ given by

$$\sin\theta_m = \frac{(m+\frac{1}{2})\lambda}{d} \qquad (m = 0, \pm1, \pm2, \ldots) \qquad \text{(minima)} \qquad (32\text{-}13)$$

It is shown in the problems that for small values of θ, for which $\theta \cong \sin\theta$, the separation distance between the mth-order maximum at a distance x_m above P_0 on the screen in Figure 32-10 and its neighboring maximum at x_{m+1} is constant and is given by

$$x_{m+1} - x_m = \frac{\lambda D}{d} \qquad (32\text{-}14)$$

This means that if we carry out an experiment with fixed values for D and d, a measurement of the distance $(x_{m+1} - x_m)$ determines in effect the wavelength λ of the radiation emitted by the source. Thus by simply measuring the distance between neighboring maxima we can determine the wavelength of the light involved. This is the first practical method for measuring the wavelength of visible light that we have come across up to this point.

As will be seen in Section 32-8 and Problem 33, because of diffraction effects the number of maxima that will appear on the screen in Figure 32-10 is severely limited. In general, only a small number of bright maxima are visible, and these appear near the center of the screen at the point P_0. The intensities of the remaining ones drop off sharply with increasing distance from P_0.

Example 32-4 For the double-slit arrangement in Figure 32-10, suppose that the slits are a distance of 0.03 mm apart, that the screen is at a distance of 50 cm from the slits, and that the wavelength of the incident light is 4000 Å.

(a) What value of θ corresponds to the first-order maximum?

(b) What value of θ corresponds to the seventh minimum?

(c) What is the spacing on the screen between successive maxima for the first few orders?

Solution

(a) Setting $m = 1$ in (32-12), we find that

$$\sin\theta_1 = \frac{4.0\times10^{-7}\text{ m}}{3.0\times10^{-5}\text{ m}} = 1.3\times10^{-2}$$

and since this angle is small it follows that

$$\sin\theta_1 \cong \theta_1 = 0.013\text{ rad}$$

(b) Setting $m = 6$ in (32-13), we find that

$$\sin\theta_6 = \left(6+\frac{1}{2}\right)\frac{\lambda}{d} = 6.5\times\frac{4.0\times10^{-7}\ \text{m}}{3.0\times10^{-5}\ \text{m}}$$
$$= 8.6\times10^{-2}$$

and hence

$$\sin\theta_6 \cong \theta_6 = 0.086\ \text{rad}$$

(c) It is apparent from the answers to (a) and (b) that for the first several orders the angles $\theta_1, \theta_2, \ldots$ are small enough that (32-14) is applicable. Hence

$$x_{m+1} - x_m = \frac{\lambda D}{d} = \frac{(4.0\times10^{-7}\ \text{m})\times(0.5\ \text{m})}{3.0\times10^{-5}\ \text{m}}$$
$$= 6.7\ \text{mm}$$

and this is large enough to be easily observable.

32-6 Fresnel's double mirror and biprism

Besides the apparatus used by Thomas Young, various other experimental setups have been suggested from time to time to produce interference effects. The purpose of this section is to describe two of these, which were originally proposed by Fresnel.

Figure 32-12a shows an apparatus known as *Fresnel's double mirror.* It consists of a small, nearly monochromatic, source S near two planar mirrors M_1 and M_2, which make a small angle α with each other. The virtual images S_1 and S_2 of the source S in the two mirrors act as two coherent sources separated by a certain distance d. Hence if any pairs of parallel rays, such as A and B, from these sources are brought together on a screen they will interfere. One important advantage of this apparatus over Young's setup in Figure 32-9 is that the distance d between S_1 and S_2 can be easily varied, within certain limits, simply by changing the angle α between the mirrors.

A second method for producing an interference pattern is by use of a

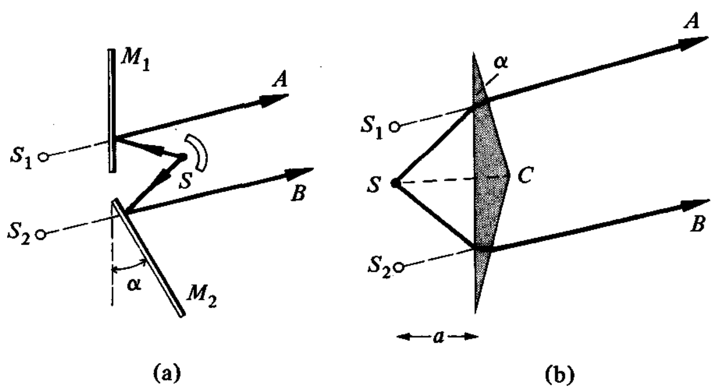


Figure 32-12

small angle prism known as a *Fresnel biprism* (see Figure 32-12b). If a point monochromatic source S is placed on the axis SC of, and at a distance a from, the prism, then when viewed from the other side, light rays appear to come from two virtual sources S_1 and S_2 located symmetrically with respect to S. In effect, then, as above, S_1 and S_2 constitute two coherent sources whose emitted radiation can be made to interfere. It is shown in the problems that for this case the distance d between S_1 and S_2 is

$$d = 2a\alpha(n-1) \tag{32-15}$$

with n the refractive index of the prism and α the angle defined in the figure. Making use of this value for d and defining the angle θ as the angle made by the parallel rays S_1A and S_2B with the axis SC of the prism, we can obtain an explicit formula for the spatial variation of the interference pattern by substitution into (32-10) and (32-11).

32-7 The diffraction grating

Consider, in Figure 32-13, plane waves of wavelength λ normally incident on a flat barrier out of which have been cut a large number N of very thin parallel slits, each separated from its nearest neighbors by a distance d. If N is sufficiently large, say of the order of 5×10^3 per centimeter of barrier, then this system is known as a *grating* or a *diffraction grating*. The number of slits per unit length of the grating is known as the *grating constant*.

As shown in the figure, suppose that on the other side of the grating there are a thin, converging lens of focal length f and a screen in the focal plane of the lens. To determine the rays which are brought to a focus at a point P on the screen, draw a line from P through the center C of the lens and let θ be the angle this ray makes with the lens axis. It follows from the properties of thin lenses that all rays which emerge from the slits and are parallel to CP will come to a focus at P. The situation here is very analogous to that analyzed in Section 32-5, where it was assumed that the screen was very far

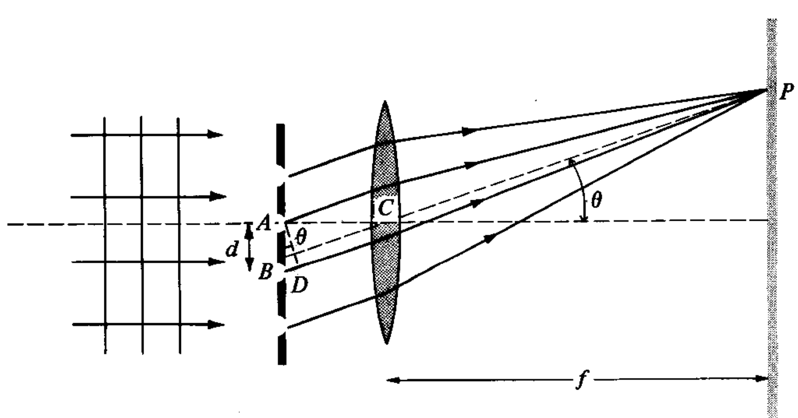


Figure 32-13

away from the slits. The function of the lens, as noted there, is simply to bring the parallel rays to a focus.

Now since the radiation incident on the grating in Figure 32-13 is a plane wave parallel to the grating, it follows that each slit may be considered to be a source of radiation that is coherent and therefore in phase with the radiation coming from the other slits. In particular, the rays that emanate from two neighboring slits A and B and are focused on the screen at P start out in phase at the points A and B, respectively. The common perpendicular AD to these two rays defines the wavefront of the wave which will be focused at P. Hence the time required for the light to travel from A through the lens to P is precisely the same as that required for the light to travel from D through the lens to P. This means that BD is the path difference Δ from the "sources" at A and B to the observer at P. Moreover, this argument can be repeated for all slits of the grating, and therefore this value for Δ is also the path difference relative to P of the radiation emanating from all neighboring slits. Reference to the figure shows that, just as for the two-slit case, this path difference Δ is

$$\Delta = d \sin \theta \tag{32-16}$$

The problem of obtaining the radiation pattern formed by a diffraction grating is thus that of calculating the radiation emitted by N coherent sources each of which has this same path difference Δ relative to its neighbor. But this has been solved in Example 32-3. Hence, substituting (32-16) into (32-9), we find that the intensity I at the point P on the screen in Figure 32-13 is

$$I = I_0 \frac{\sin^2 [(N\pi d \sin \theta)/\lambda]}{\sin^2 [(\pi d \sin \theta)/\lambda]} \tag{32-17}$$

The constant I_0 may be thought of as the intensity that would be observed if only one slit were open. It is left as an exercise to show that for the special case $N = 2$, this is precisely equivalent to (32-11), as it must be.

Figure 32-14 shows a plot of the formula for I in (32-17) as a function of the quantity $(\pi d \sin \theta)/\lambda$ for a value of $N \geqslant 10$. (For common gratings a more typical value is 5×10^3.) Note that there is a series of very sharp peaks,

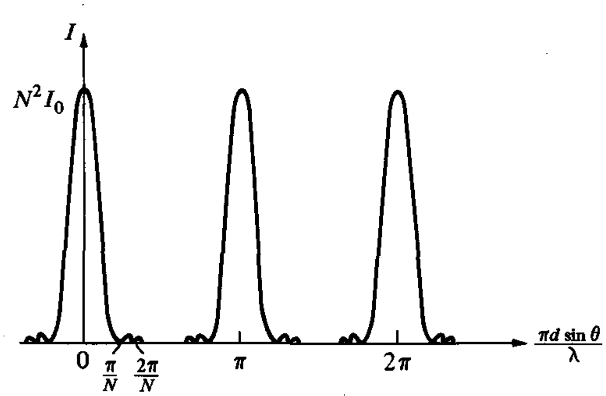


Figure 32-14

each of height N^2I_0 and each with a width of approximately $2\pi/N$. The separation distance between peaks corresponds, according to (32-16), to the path differences Δ being an integral multiple of λ. On either side of these sharp peaks there are additional maxima and minima, but these are generally of the order I_0 and thus of no significance except for the lowest values of N.

In terms of the angle θ, which characterizes the point P on the screen in Figure 32-13, we may summarize the relation in (32-17) in the following way. For a given wavelength λ, the observed intensity I achieves its maximum of N^2I_0 at those values θ_m given by

$$d \sin \theta_m = m\lambda \qquad (m = 0, 1, 2, \ldots) \tag{32-18}$$

For at these values the ratio on the right-hand side of (32-17) assumes the form

$$\frac{\sin^2 Nm\pi}{\sin^2 m\pi}$$

which is indeterminate since N and m are both integers, and thus both the numerator and the denominator vanish. However, this ratio may be evaluated by a limiting process and as shown in the problems the result may be expressed as

$$\lim_{x \to m} \frac{\sin^2 N\pi x}{\sin^2 \pi x} = N^2$$

for m an integer. The line corresponding to $m = 0$, is known as the *central* maximum, and is observed at $\theta = 0$ for *all wavelengths*. The line corresponding to $m = 1$ is called the *first-order* maximum, that at $m = 2$ is the *second-order* maximum, and so forth. The associated angles θ_1 and θ_2 may be obtained from (32-18) by the substitution of $m = 1$ and $m = 2$, respectively. Note that since $|\sin \theta| \leq 1$, the maximum order observable is the greatest integer not exceeding the ratio d/λ.

Diffraction gratings play an extremely important role in optical spectroscopy. According to (32-18), if the spacing d for a grating is given, the measurement of the wavelength of a line is reduced to that of making a measurement of an appropriate angle. A variety of optical gratings is available commercially, including both reflection and transmission gratings.

Although it has been implicitly assumed in all of the above that the radiation of interest was visible light, it should be noted that interference effects can be produced for electromagnetic radiation of any wavelength. Indeed, in 1913 Max von Laue (1879–1960) suggested that a crystalline solid might act as a (reflection) grating for X rays. As we know today, the wavelengths of X rays are of the order of 10^{-7} cm to 10^{-8} cm, and this is comparable to the interatomic spacing expected in crystals. The first experiments along these lines—which are known today as *X-ray diffraction experiments*—were carried out shortly thereafter by Friedrich and Kipping. They confirmed not only the fact that X rays form a part of the electromagnetic spectrum, but also that the atoms in some solids are arranged

in regular, crystal-like arrays. This phenomenon of X-ray diffraction is still used today as an important tool in studying the properties of both organic and inorganic crystals.

Example 32-5 Light of wavelength 6000 Å falls on a diffraction grating having 5000 lines/cm.

(a) Calculate the grating spacing.

(b) At what angle θ_1 does the first-order line occur?

(c) If a lens of focal length 12 cm is used to focus the light on a screen, how far is it from the central maximum to the first order?

Solution

(a) Since there are 5000 lines/cm, it follows that the distance d between the lines is

$$d = \frac{1}{5 \times 10^3/\text{cm}} = 2.0 \times 10^{-4}\ \text{cm}$$

(b) Use of the value $m = 1$ in (32-18) yields for θ_1

$$\sin\theta_1 = \frac{\lambda}{d} = \frac{6.0 \times 10^{-7}\ \text{m}}{2.0 \times 10^{-6}\ \text{m}} = 0.30$$

whence

$$\theta_1 \cong 17.5°$$

(c) Reference to Figure 32-13 shows that

$$\tan\theta_1 = \frac{x}{f}$$

where x is the distance along the screen to the first-order maximum from the central maximum and f is the focal length of the lens. Using the known values for θ_1 and f we find that

$$\begin{aligned} x = f\tan\theta_1 &= (12\ \text{cm}) \times \tan 17.5° \\ &= (12\ \text{cm}) \times 0.31 \\ &= 3.7\ \text{cm} \end{aligned}$$

Example 32-6 Light of wavelength 4500 Å is incident on a grating of 6000 lines/cm. What orders are observed?

Solution Solving (32-18) for m we find that

$$\begin{aligned} m &= (\sin\theta_m)\frac{d}{\lambda} = \sin\theta_m \frac{1.67 \times 10^{-4}\ \text{cm}}{4.5 \times 10^{-5}\ \text{cm}} \\ &= 3.7 \sin\theta_m \end{aligned}$$

Since $|\sin\theta| \leq 1$, it follows that only for the values $m = 0, 1, 2,$ and 3 does there exist an angle θ_m satisfying this relation. Hence, besides the central maximum, only the first-, second-, and third-order maxima will be observed.

32-8 Fraunhofer diffraction

On passing through an aperture in a barrier, a beam of light will spread out, to some extent, into the region where we would normally expect to see a shadow. As noted earlier, this is an example of the diffraction of light, or the property of light of bending around sharp edges. The distinction between diffraction and interference is not very precise and for our purposes a diffraction pattern may be thought of as the pattern formed by the interference of a very large number of coherent beams.

Historically, the two classes of diffraction phenomena that have been analyzed in depth are *Fresnel diffraction* and *Fraunhofer diffraction.* Figure 32-15 shows how a Fraunhofer diffraction pattern may be produced. Light from a line (or point) source S is rendered parallel by a positive lens L_1 and the resultant plane waves are incident on a barrier AB out of which has been cut a slit of width b. After the light emerges from the slit, the resultant rays are brought to a focus on a screen or a photographic plate CD, which is located in the focal plane of a second positive lens L_2. Note that waves which arrive at the barrier AB, as well as those that arrive at the screen CD, are plane waves. This feature is a main characteristic of Fraunhofer diffraction. If either of the lenses is removed so that waves that approach either the slit or the screen are spherical or cylindrical waves, then the observed pattern is said to be due to Fresnel diffraction. In Fraunhofer diffraction the distances from the source to the slit and from the slit to the screen are both effectively infinite. Fresnel diffraction takes place if either or both of these distances are finite. For reasons of simplicity, we shall consider here only Fraunhofer diffraction by a single slit.

Consider, in Figure 32-16, plane waves incident on a slit of width b and suppose that the resultant diffracted rays are brought to a focus on a screen by a lens L. In accordance with the Huygens principle, each point of the slit may be thought of as being a source of outgoing spherical waves. Moreover, as set up, these sources are coherent. To calculate the actual intensity $I(\theta)$ at

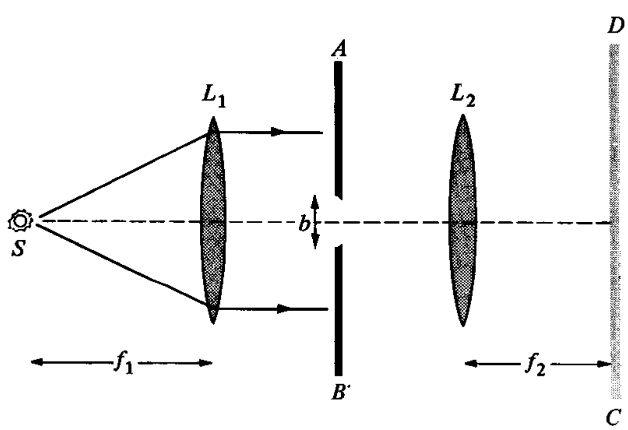


Figure 32-15

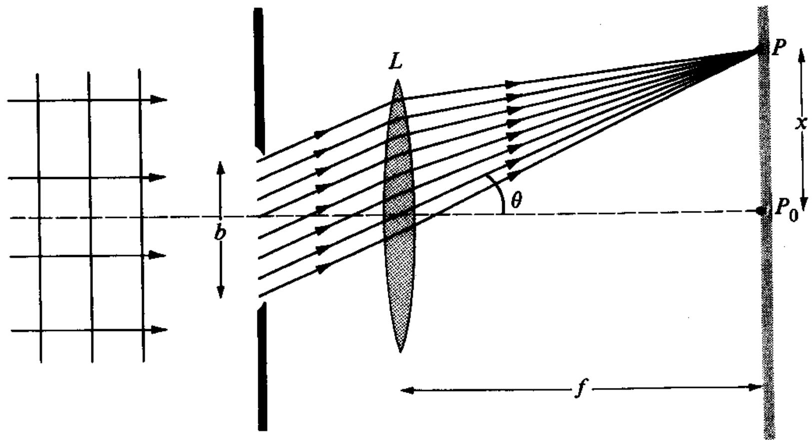


Figure 32-16

the point P on the screen, we may suppose, therefore, that in the region of the slit there are N equally spaced and coherent point (or line) sources and that we are interested in the resultant interference pattern for very large N. The distance d between these sources is given in terms of the slit width b by

$$d = \frac{b}{N} \tag{32-19}$$

and thus according to (32-17), the intensity $I_N(\theta)$ for N sources may be expressed by

$$I_N = \frac{I_m}{N^2} \frac{\sin^2 [(N\pi d \sin \theta)/\lambda]}{\sin^2 [(\pi d \sin \theta)/\lambda]}$$

where $I_m \equiv I_0 N^2$ represents the maximum intensity of the principle maximum (Figure 32-14). Applying this formula to the situation in Figure 32-16, we obtain by use of (32-19)

$$I_N = \frac{I_m}{N^2} \frac{\sin^2 [(\pi b \sin \theta)/\lambda]}{\sin^2 [(\pi b \sin \theta)/N\lambda]}$$

Passing to the limit $N \to \infty$, we find that the intensity $I(\theta)$ for the Fraunhofer diffraction pattern is given by

$$I(\theta) = I_m \frac{\sin^2 \beta}{\beta^2} \tag{32-20}$$

where $\beta \equiv \beta(\theta)$ is defined by

$$\beta = \frac{\pi b \sin \theta}{\lambda} \tag{32-21}$$

and we have used the fact that as N becomes very large, $N^2 \sin^2 (x/N) \to x^2$.

Figure 32-17 shows a plot of $I(\theta)$ as a function of $\beta = (\pi b \sin \theta)/\lambda$. Since the ratio $(\sin x)/x$ approaches unity as x tends to zero, it follows from (32-20) that at the value $\beta = 0$, $I(\theta)$ achieves its maximum value of I_m. The intensity function $I(\theta)$ vanishes whenever β assumes an integral multiple value of π. Hence, since $I(\theta)$ is nonnegative, it follows that there are

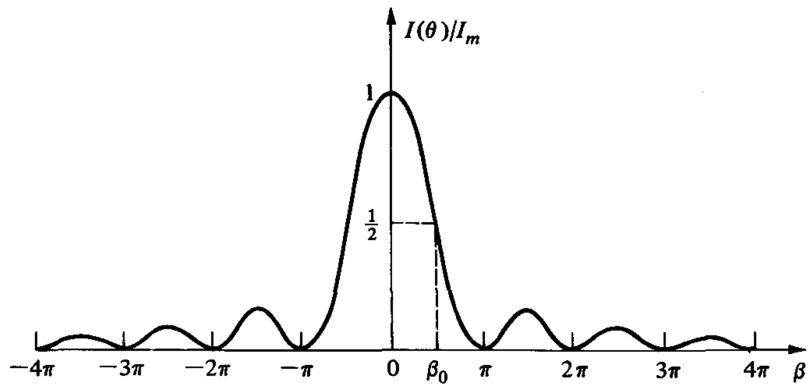


Figure 32-17

additional maxima between each pair of these zeros. To a rough approximation we may assume that these maxima lie midway between these zeros—that is, at the values $\beta \cong 3\pi/2, 5\pi/2, 7\pi/2, \ldots$. Substituting these values for β into (32-20), we find that at these secondary maxima $I(\theta)/I_m$ assumes the values $(2/3\pi)^2, (2/5\pi)^2, (2/7\,\pi)^2, \ldots$; or, equivalently, the values 0.045, 0.016, 0.0083, ... Thus the large central maximum, for which the ratio $I(\theta)/I_m$ is unity, dominates the entire diffraction pattern! Nevertheless, the first maximum at $\beta \cong 3/2\pi$ has an intensity of about 4.5 percent of that of the central maximum and is thus frequently visible.

In order to describe the width of the central maximum in Figure 32-17, it is convenient to define the parameter β_0 as the value of β for which the intensity has dropped to half of its maximum value. Substituting the values $\beta = \beta_0$ and $I = I_m/2$ into (32-20) yields

$$\sin^2 \beta_0 = \frac{\beta_0^2}{2} \tag{32-22}$$

so that

$$\beta_0 \cong 1.4 \text{ rad} = 80°$$

The associated angle θ_0 may be found by substitution into (32-21).

A second angle of interest in this connection is the angle θ_1, defined as the position on the screen where the intensity of the central maximum has dropped to zero. Since the value for β associated with θ_1 is π, we find by substitution into (32-21) that

$$\sin \theta_1 = \frac{\lambda}{b} \tag{32-23}$$

Physically, we may understand the significance of θ_1 in the following way. If light were not a wave, then the light coming through the aperture in Figure 32-16 would obey the laws of geometrical optics and come to a focus at the point P_0 on the screen. However, light *is* a wave phenomenon. Hence this point image is diffracted out until it has an angular diameter given approximately by twice the angle θ_1 in (32-23). In the limit of very large b, the width of the central maximum shrinks to zero, corresponding to a point image.

That is, for an aperture size very large compared to λ, the laws of geometrical optics prevail.

At the other extreme, where the slit width b is very small compared to λ, the intensity on the screen appears to be uniform. This is the situation shown in Figure 32-2. In the analysis of the two-slit diffraction pattern in Figure 32-9 and in the corresponding analysis for the diffraction grating in Section 32-7, it was implicitly assumed that $b \ll \lambda$. For the case $b \sim \lambda$ a somewhat more complex analysis is required. The result of such an analysis for the two-slit case is given in Problem 33 and is that $I(\theta)$ consists of two factors: one of these is the single-slit Fraunhofer pattern $(\sin^2\beta)/\beta^2$ in (32-20); the other is the two-slit distribution $\cos^2[(\pi d \sin\theta)/\lambda]$ in (32-11). In the analysis carried out in Section 32-5, it was implicitly assumed that the slits were very narrow ($b \ll \lambda$), so that $(\sin^2\beta)/\beta^2 \cong 1$ and the first of these factors could be neglected. However, for large enough angles this is no longer true. It is for this reason that the two-slit pattern observed on the screen in Figure 32-10 extends out only to small angles, for which diffraction effects are negligible.

Example 32-7 A slit of width 0.02 mm is illuminated normally by plane monochromatic light waves of wavelength 6000 Å. A thin, positive lens of focal length 50 cm is used to project the diffraction pattern onto a screen (see Figure 32-16).

(a) What is the angular separation from the center of the principal maximum to the first minimum?

(b) At what value of θ has the intensity of the central maximum decreased to half its maximum value?

(c) How wide is the central maximum as observed on the screen?

Solution

(a) The required value of θ corresponds to $\beta = \pi$, and is given by the value θ_1 in (32-23). Substituting the known values for λ and b, we find that

$$\sin\theta_1 = \frac{\lambda}{b} = \frac{6.0 \times 10^{-5}\ \text{cm}}{2.0 \times 10^{-3}\ \text{cm}} = 0.03$$

Thus, we may replace $\sin\theta_1$ by θ_1 and conclude that

$$\theta_1 = 0.03\ \text{rad} = 1.7^\circ$$

(b) This angle is the angle θ_0 in (32-22). Hence by use of (32-21) it follows that

$$\sin\theta_0 \cong \theta_0 = \frac{1.4\lambda}{\pi b} = \frac{1.4}{\pi} \times 0.03 = 0.013\ \text{rad}$$

(c) Reference to Figure 32-16 shows that the distance x_1 to the screen associated with the angle θ_1 is

$$x_1 = f \tan\theta_1 \cong f\theta_1 = (50\ \text{cm}) \times 0.03\ \text{rad} = 1.5\ \text{cm}$$

where $f = 50$ cm is the focal length of the lens and we have used the result of (a). The width of the central maximum is $2x_1$, or 3 cm.

32-9 Resolving power

Consider, in Figure 32-18, two *incoherent* slit sources S_1 and S_2, separated by a small angular distance α and emitting light waves that come to a focus on a screen *CD* after going through an aperture of width *b*. Assuming that the distances between the sources and the aperture and between the aperture and the screen are effectively infinite, we expect to find on the screen two Fraunhofer diffraction patterns with centers separated by the same angle α. By contrast to the laws of geometrical optics according to which there should appear two points of light on the screen, one at P_1 and the other at P_2, we find instead around each of these two points P_1 and P_2 an illuminated region associated with the diffraction patterns produced by each of the sources. If the angular separation α between the sources is very small, it may not be possible to ascertain the existence of two distinct sources. The purpose of this section is to see under what circumstances it is possible to distinguish clearly between two sources very close together.

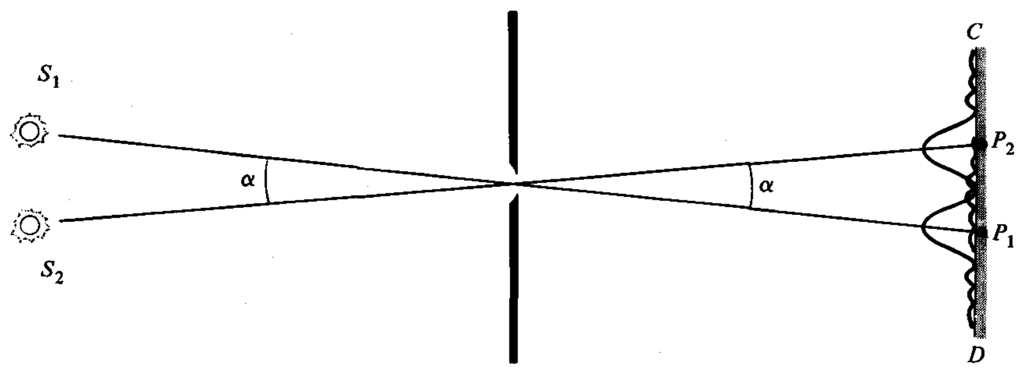


Figure 32-18

Figure 32-19 shows the various classes of diffraction patterns that can arise from two sources separated by a very small angle α. In each case, the dashed lines represent the separate diffraction patterns of the two sources, and the solid line, which represents their sum, is the observed intensity. Recall that the sources are assumed to be incoherent. The angle θ_1, as defined in (32-23), stands for the angular distance from the center of the principal maximum of a source to the point for which the intensity vanishes. For the sake of simplicity let us assume that the maximum intensities of the two sources are the same, and that they are monochromatic and radiating at the same wavelength λ. Figure 32-19a shows the case when the angle α between the sources is much larger than θ_1. Here the two maxima are clearly separated and we have no trouble concluding that two radiators are emitting. Similarly, Figure 32-19b shows the case in which $\alpha = 1.5\theta_1$ and, again, even though there is some overlap of the principal diffraction maxima, it is still easy to see that there exist two distinct sources. Finally, Figures 32-19c and

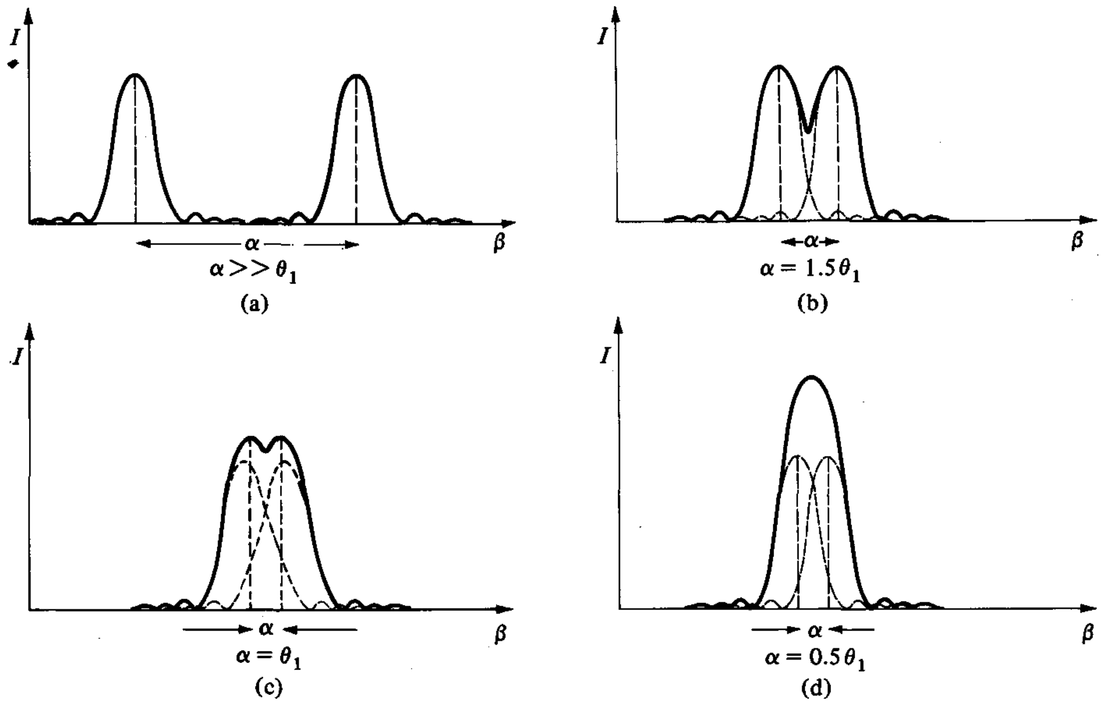


Figure 32-19

d show the cases in which $\alpha = \theta_1$ and $\alpha = 0.5\theta_1$, respectively. Although it is possible by careful observation to resolve the two sources in the former case, this is certainly not so in the latter. Hence we see that if two sources are separated by an angular distance less than θ_1, then for the given wavelength and aperture size it is no longer possible to resolve the two images.

A study of the patterns shown in Figure 32-19 leads one to the conclusion that the situation in Figure 32-19c represents the smallest separation angle α for which the images can still be resolved. It was decided, somewhat arbitrarily perhaps, by Lord Rayleigh that this angle α_1, which is given by

$$\alpha_1 = \theta_1 = \frac{\lambda}{b} \tag{32-24}$$

where the second equality follows by use of (32-23) with $\theta_1 \ll 1$, should be the *criterion* for the resolvability of the two diffraction patterns. This choice for the angle of separation α_1 between the two images is known as *Rayleigh's criterion,* and defines the *resolving power* for the aperture b. In practical terms, what this means is that if the separation angle α between two sources is less than the value α_1 given in (32-24) then they cannot be resolved by use of an aperture of the given size b. Conversely, if the separation angle between two sources is greater than λ/b, then they are resolvable.

In the above discussion we have made use of (32-23) and thus have assumed that the sources and the aperture were long slits. For most cases of

practical interest we are concerned with point sources, such as two stars or the opposite edges of an amoeba, and with circular apertures, such as the objective lens of a telescope or a microscope. To apply Rayleigh's criterion to these cases requires a knowledge of the diffraction pattern produced by a circular aperture. The results of this calculation—which was first carried out by Sir George Airy (1801–1892)—are surprisingly simple and are given by

$$\alpha_1 = 1.22\frac{\lambda}{b} \tag{32-25}$$

That is, the minimum angle of separation α_1 between two point sources that radiate at the wavelength λ and which can be still resolved by use of a circular aperture of diameter b is given by $1.22\lambda/b$. So except for the factor 1.22, this formula for the resolving power of a circular aperture is the same as (32-24) for a slit.

Example 32-8 A "spy satellite" orbits the earth at a height of 100 miles. What is the minimum diameter b of the objective lens in a telescope that must be used to resolve the marked yard lines on a football field? Assume $\lambda = 5500$ Å and that the limitation is due to diffraction.

Solution The angular separation α between two lines separated by a distance of 5 yards and as seen 100 miles away is

$$\alpha_1 = \frac{5 \text{ yards}}{100 \text{ miles}} = \frac{5 \text{ yards}}{100 \text{ miles} \times 1760 \text{ yards/mile}}$$
$$= 2.8 \times 10^{-5} \text{ rad}$$

Substituting this value into (32-25) and solving for b we find that

$$b = \frac{1.22\lambda}{\alpha_1} = \frac{1.22 \times 5.5 \times 10^{-7} \text{ m}}{2.8 \times 10^{-5} \text{ rad}}$$
$$= 2.4 \text{ cm}$$

Example 32-9 Assuming that the pupil of the human eye has a diameter of 4.0 mm, what is the minimum separation distance y between a person and two small, red objects ($\lambda = 6500$ Å) separated from each other by a distance of 20 cm. (See Figure 32-20, which is *not* to scale!)

Solution According to the given data, the minimum angle α that can still be resolved is $20 \text{ cm}/y$. Equating this to $1.22\lambda/b$ in accordance with (32-25) and solving for y, we obtain

$$y = \frac{20 \text{ cm}}{1.22\lambda/b} = \frac{(20 \text{ cm}) \times (4.0 \times 10^{-1} \text{ cm})}{1.22 \times 6.5 \times 10^{-5} \text{ cm}} = 1.0 \text{ km}$$

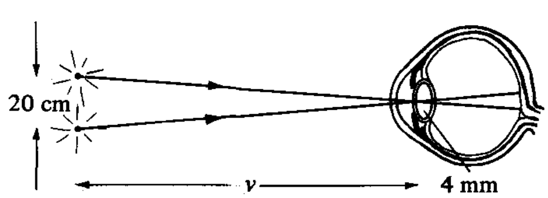


Figure 32-20

32-10 Interference by thin films

A source of illumination that is *not* a point source, and which may therefore be thought of as consisting of a large number of *incoherent* point sources, is known as an *extended* or a *diffuse* source. If light from such a source is reflected from a thin, transparent film—a thin layer of oil or the membrane of a soap bubble, for example—then under certain conditions interference effects will be observed. These effects manifest themselves frequently by the appearance of bright-colored bands across the surface of the film. The purpose of this section is to show how this phenomenon can be understood in terms of an interference effect.

Consider, in Figure 32-21, a thin film of material of thickness d and of refractive index n and illuminated by an extended monochromatic source of wavelength λ. To see how it is possible for the observer O to see interference effects, let us focus on a typical ray AB, which is emitted from a point A on the source and in a direction so that after reflection from the film it enters the observer's eye. Since the source is extended, light rays from various other points of the source will enter the observer's eye as well, but for the moment let us consider only that associated with the ray AB. As shown in the figure, not all of the light in the incident ray is reflected at the point B on the upper surface of the film. Some of it is refracted and enters the film along the direction BC. On arriving at point C, part of this refracted ray BC is reflected upward along CD, and the remainder is refracted out of the underside of the film. The ray CD, on striking the upper surface at the point D, will also be partially refracted out of the film and in a direction parallel to the ray BO. Finally, since both of the rays BO and DO, which enter the observer's eye, originally started out as a single ray at the point B, it follows that they are coherent and that a definite phase relationship exists between them. Hence, when these rays come together on a screen (the retina of the observer's eye, in this case), interference effects may be observed.

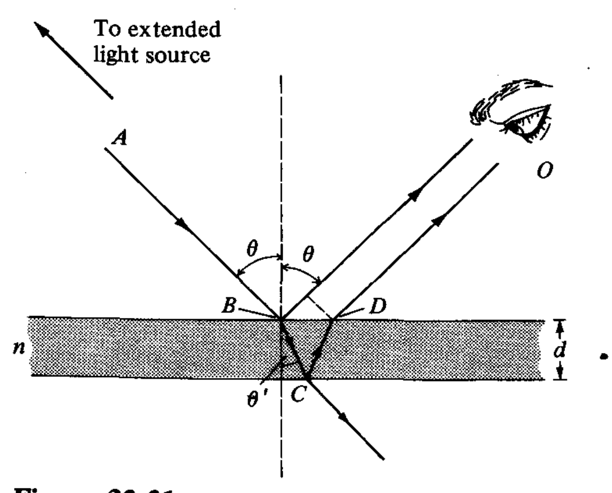


Figure 32-21

To simplify the analysis, let us confine ourselves in all of the following to observations near normal incidence. This means that $\theta \cong n\theta' \cong 0$ so that the distances $\overline{BC}$ and $\overline{CD}$ have a length d and the distance $\overline{BD}$ is negligible compared to the thickness d of the film. It follows that under this circumstance the path difference Δ between the rays BO and $BCDO$, which start out in phase at B, is

$$\Delta = 2d$$

We might expect, therefore, that constructive or destructive interference would be observed if this path difference $2d$ were an integral or a half-integral multiple of a wavelength, respectively. However, this is *not* so. Instead, as will be established below, maxima and minima in intensity occur in accordance with the formulas:

$$\begin{aligned} 2nd &= (m+\tfrac{1}{2})\lambda \qquad &\text{(maxima)} \\ 2nd &= m\lambda \qquad &\text{(minima)} \end{aligned} \tag{32-26}$$

for m a nonnegative integer. These formulas differ in two ways from what one might expect based on our previous studies of interference: (1) by the appearance of the index of refraction n of the film on the left-hand side; and (2) by the interchange of the roles of maxima and minima in terms of the number of wavelengths in the path difference. The quantity $2nd \equiv n\Delta$ is also known as the *optical path difference.*

The necessity for the factor n in (32-26) can be seen in the following way. The wavelength λ in these formulas refers to the free-space wavelength of the light and not to its wavelength λ/n in the medium; see (29-24). Since the path difference of the interfering rays in Figure 32-21 is the extra distance, $\overline{BC}+\overline{CD}=2d$, traveled by the ray $BCDO$ through the film, it follows that the wavelength λ/n in *the medium* is the one relevant for determining whether or not interference effects take place. Thus the reason for the factor n in (32-26) is so that the wavelength λ in these formulas refers to the free-space value.

With regard to the reversals of the roles of maxima and minima in (32-26), we may argue as follows. According to the first of the Fresnel equations in (30-17) the amplitude E_0'' of an electromagnetic wave reflected at normal incidence, at the interface between two media, will undergo a reversal of direction if $n_1 < n_2$. In other words, light incident from the lower index side of the interface between two media will undergo a phase change of 180° on reflection. According to the general relation between phase difference δ and path difference Δ in (32-7), a phase difference of 180°, or π radians, corresponds to a path difference of a half of a wavelength. Therefore, if the path difference $2d$ in Figure 32-21 is an integral number of wavelengths, the interference is destructive. Correspondingly, constructive interference will be associated with a path difference $2d$ equal to an odd half integral multiple of a wavelength λ/n. It is for this reason then that the roles of maxima and minima are reversed in (32-26). Note that even if the film in Figure 32-21 is in

a medium of refractive index n' ($>n$), the relations in (32-26) are still applicable. According to (30-17), for this case a phase change of 180° occurs for the ray reflected at the lower surface, whereas at the upper surface the incident and the reflected rays remain in phase.

32-11 Observations of thin-film interference

In this section we apply (32-26) to several physical situations.

Consider first, in Figure 32-22a, light from a *monochromatic* extended source normally incident on two glass plates, which touch at one end and are separated by a small spacer, such as a piece of paper, at the other. In effect, we have a thin film of air whose thickness varies along the plates. It follows from (32-26) that the observer of the reflected rays will see a series of alternating dark and bright bands across the plate. See Figure 32-22b. Along the edge where the plates are in contact, $d = 0$. Hence, regardless of λ, (32-26) implies that destructive interference will be observed here. Further, since $n = 1$ for air, it follows, by setting $m = 0$ in (32-26), that for $d = \lambda/4$ a bright band will be seen on the plate. As shown in Figure 32-22b, bright bands will appear at film thicknesses of $3\lambda/4$, $5\lambda/4, \ldots$, and these will be separated by dark bands with centers corresponding to film thicknesses of $d = \lambda/2, \lambda, 3\lambda/2, \ldots$.

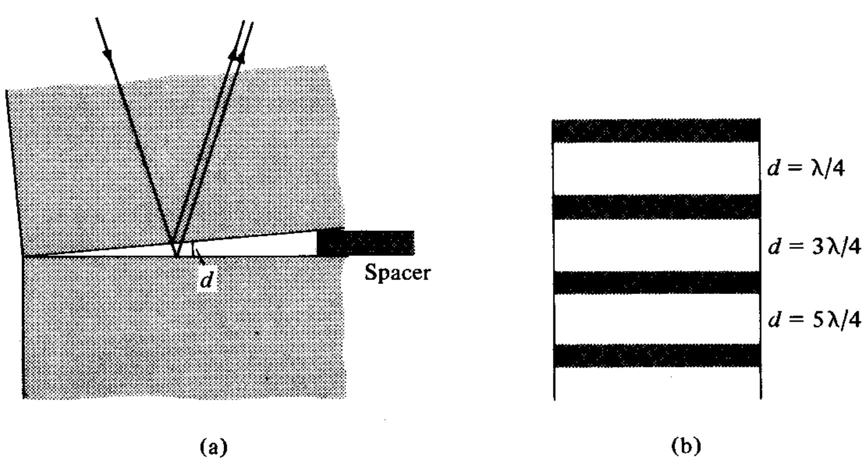


Figure 32-22

If the source is not monochromatic, then bands of light of other wavelengths may also appear on the plate. For example, if the source is sunlight, with its full spectrum of colors, the observer will see a sequence of bright bands with continuously varying colors from red to violet.

A second apparatus for observing interference in an air film is shown in Figure 32-23. Light from an extended source S (not shown) is normally incident on the flat side of a plano-convex lens B whose spherical face lies on a flat glass plate A. The air film between A and B then produces circular

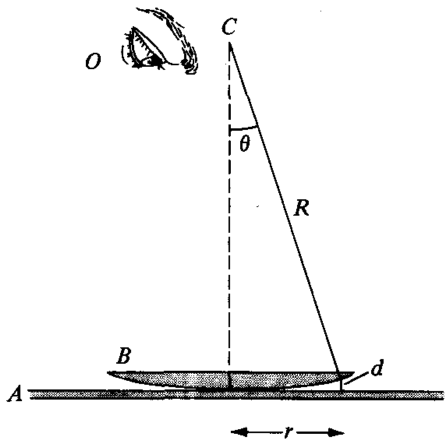


Figure 32-23

interference fringes, which will be seen by the observer O. These particular fringes are known as *Newton's rings* and were first studied in detail by Isaac Newton. It is ironic that these rings, which were really the first experimental evidence for the wave nature of light, should be credited to the creator and defender of the corpuscular theory.

To obtain a quantitative relation between the radius and wavelength of a given ring, let us consider the apparatus in more detail. If r is the radius of a given ring of order m, and d the corresponding thickness of the air film, then for small angles

$$d = R - R\cos\theta \cong R - R\left(1 - \frac{\theta^2}{2}\right) = \frac{R\theta^2}{2}$$
$$= \frac{r^2}{2R}$$

since $\cos\theta \cong 1 - (\theta^2/2)$ and, according to the figure, $\theta \cong \tan\theta = r/R$. Substituting this form for the thickness of the air film into (32-26) and using the value $n = 1$ for air, we find for the radius r_m of the $(m+1)$th bright ring

$$r_m = \left[R\lambda\left(m + \frac{1}{2}\right)\right]^{1/2} \qquad m = 0, 1, 2, \ldots \tag{32-27}$$

Because of the fact that there is a phase change associated with the reflection from the lower glass plate, the spot in the middle of the plate is dark. That is, destructive interference takes place at the central point. This feature is clearly illustrated in Figure 32-24, which is an actual photograph of the fringes formed by this method.

Finally, let us consider in Figure 32-25 a precision instrument known as the *Michelson interferometer*, which also produces interference fringes when exposed to an extended source. Light from a diffuse source S on arriving at a half-silvered plate P_1 is broken up into a reflected beam (1) and a transmitted beam (2) of approximately equal intensity. Beam 1 travels upward toward a highly polished mirror M_1, from which it is reflected back through P_1 into the eye of the observer at O. The second beam is directed

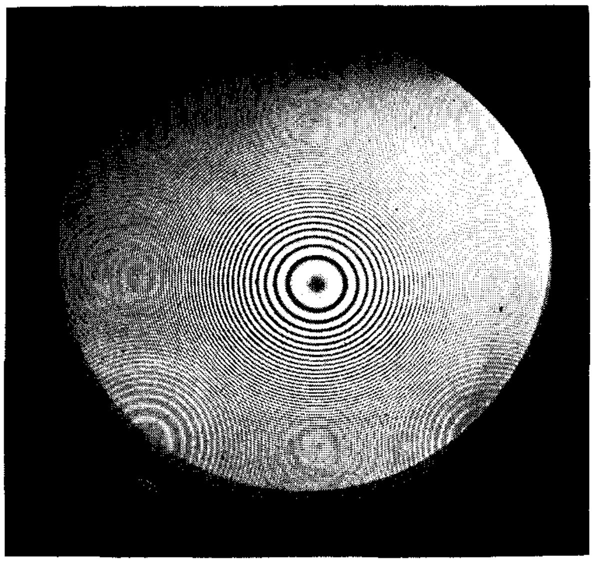


Figure 32-24 *Newton's rings produced by use of a monochromatic source. (Courtesy Bausch and Lomb, Scientific Optical Products Division.)*

first through a compensating plate P_2 so if $l_1 = l_2$, then the optical paths for the two beams will be the same. It is then reflected from a second mirror M_2, back through P_2, and finally into the observer's eye after being reflected by P_1. The observer O thus sees two images of the extended source S; one of these comes from its reflection in M_1 and the other in M_2. Just as for the situation shown in Figure 32-21, the rays that appear to come from these two images will interfere constructively or destructively depending on the number of wavelengths in the optical path difference between rays (1) and (2). For example, if the mirrors M_1 and M_2 are not precisely perpendicular to each other, then in effect we have the situation in Figure 32-22 and,

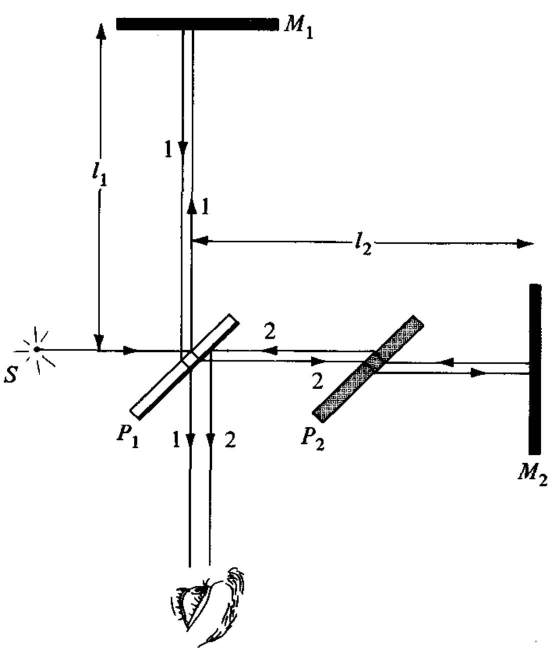


Figure 32-25

consistent with experiment, the observer sees a set of parallel, bright fringes. If the two mirrors are precisely perpendicular, then the observer sees a series of concentric circular fringes, reminiscent of Newton's rings.

The interferometer has played, and continues to play, an important part in precision measurements of various physical quantities. It was originally designed by Michelson in 1887 to measure the velocity of the earth through the "aether." The negative outcome of this Michelson–Morley experiment, as it is known, showed that the velocity of light is independent of the relative motion between the source and the observer. The significance of this fact was not fully appreciated until later when Einstein enunciated the special theory of relativity. Subsequently, the interferometer was used to measure the diameters of selected *red-giant* stars. More recently, the interferometer has been used to define the SI unit of length of the meter as the distance equal to 1,650,763.73 wavelengths of the bright orange-red line in the spectrum of krypton-86. The precision possible with this instrument is truly remarkable.

Example 32-10 Newton's rings are observed by use of light of wavelength 6000 Å and a plano-convex lens with a radius of curvature of 25 cm. What is the radius of the first and the twentieth ring?

Solution Making use of the given values $R = 25$ cm and $\lambda = 6.0 \times 10^{-5}$ cm, we find by setting $m = 0$ in (32-27) that the radius r_1 of the first ring is

$$r_1 = \left[R\lambda\left(m + \frac{1}{2}\right)\right]^{1/2} = \left[25\text{ cm} \times 6.0 \times 10^{-5}\text{ cm} \times \frac{1}{2}\right]^{1/2}$$
$$= 0.27\text{ mm}$$

while that for the twentieth ring is found by setting $m = 19$

$$r_{20} = [25\text{ cm} \times 6.0 \times 10^{-5}\text{ cm} \times 19.5]^{1/2}$$
$$= 1.7\text{ mm}$$

Example 32-11 In Figure 32-23 suppose that plate A has a refractive index n_A, plate B a refractive index n_B, and the intervening space is filled up with a fluid of refractive index n_C. If

$$n_A < n_C < n_B \quad \text{or} \quad n_A > n_C > n_B$$

what is the formula analogous to (32-27) for the radius of the bright rings?

Solution The derivation would go through just as above, but now there is *no* relative phase change on reflection. That is, at each reflection both rays suffer either no phase change or else both undergo a phase change of 180°. In either case the relative phase change is zero. Thus in place of the first of (32-26), the condition for a maximum becomes

$$2n_C d = m\lambda$$

Proceeding as in the derivation of (32-27), we conclude that rings occur now at the values of r given by

$$r_m = \left[\frac{R\lambda m}{n_C}\right]^{1/2} \qquad m = 0, 1, 2, \ldots$$

In particular, this means that a bright spot appears at the center of the plate under these circumstances!

32-12 Summary of important formulas

If, as in Figure 32-13, a plane monochromatic wave is incident on a barrier out of which have been cut N very narrow slits, the intensity $I(\theta)$ observed on a distant screen is

$$I(\theta) = I_0 \frac{\sin^2 [(N\pi d \sin\theta)/\lambda]}{\sin^2 [(\pi d \sin\theta)/\lambda]} \tag{32-17}$$

where d is the spacing between the slits, λ the wavelength, and I_0 the intensity that would be observed if only one slit were present. For the special case $N = 2$ this reduces to

$$I(\theta) = 4I_0 \cos^2 \left[\frac{\pi d \sin\theta}{\lambda}\right] \tag{32-11}$$

In both (32-17) and (32-11) it is assumed that the width of each slit is very small compared to λ.

If plane monochromatic light of wavelength λ is incident on a slit of width b, and if we define the angle θ as in Figure 32-16, then the observed intensity $I(\theta)$ at the point θ is

$$I(\theta) = I_m \frac{\sin^2 \beta}{\beta^2} \tag{32-20}$$

where the parameter β is defined by

$$\beta = \frac{\pi b \sin\theta}{\lambda} \tag{32-21}$$

and where I_m is the intensity of the central maximum (Figure 32-17).

If two distant and *incoherent* point objects, such as two stars, are viewed through a circular aperture of size b, and if the minimum angle between the sources exceeds the angle α_1, given by

$$\alpha_1 = 1.22 \frac{\lambda}{b} \tag{32-25}$$

then the sources can be resolved.

If monochromatic light from a diffuse or extended source is reflected by a thin film of refractive index n and of thickness d, then at normal incidence

there will be maxima and minima in the observed intensity for

$$\begin{aligned} 2nd &= (m + \tfrac{1}{2})\lambda \qquad \text{(maxima)} \\ 2nd &= m\lambda \qquad \text{(minima)} \end{aligned} \tag{32-26}$$

for m a nonnegative integer.

QUESTIONS

1. Define or describe briefly what is meant by the following: **(a)** interference; **(b)** Huygens construction; **(c)** incoherent sources; **(d)** destructive interference; and **(e)** Fresnel diffraction.
2. The light emitted from two very close stars is brought to a focus on a screen by a lens. Will an interference pattern appear on the screen? Explain.
3. It is often stated that a diffraction pattern is the result of the interference of the radiation emitted by a very large number of coherent sources. Discuss the meaning of this statement in terms of Huygens principle.
4. Explain why two coherent beams of light of the same amplitude must be traveling in parallel—or nearly parallel—paths in order for complete destructive interference to be observed.
5. In Figure 32-13, why is the path difference from A to P the same as that from D to P regardless of the detailed structure of the lens? (*Hint*: According to Fermat's principle, all rays that emanate from a point source spend the same amount of time traveling to a point of focus after going through a lens.)
6. An interference pattern is formed by two coherent light beams, one of which has twice the intensity of the other. In what way does this pattern differ from the one that would be formed if the beams were of the same strength?
7. Explain, in terms of a microscopic picture, why the light rays emitted from different points of an ordinary source are incoherent.
8. A series of equidistant interference maxima are observed in a two-slit experiment, such as in Figure 32-9. If the separation distance d between the slits is doubled, what happens to the spacing between the fringes on the screen? If light of a smaller wavelength is used what happens to the distance between the fringes?
9. Suppose that the angle α of the Fresnel double mirror in Figure 32-12a is halved. What happens to the distance between the virtual sources S_1 and S_2? What happens to the separation distance between neighboring fringes on the screen?
10. Explain why the vertex angle C of the Fresnel biprism in Figure 32-12b must be very nearly 180° in order for fringes to be produced.
11. A diffraction grating has a certain number of lines per unit length. How do you find, from this datum, the values for the parameters d and N to be used in connection with (32-17)? If you cannot obtain values for both d and N, what additional data do you require?
12. Explain why it is desirable to have the rulings of a diffraction grating as close together as possible. Why is it desirable also to have a large value for the parameter N?
13. How do you account for the existence of an upper limit to the order m of a principal maximum that can be

obtained by use of a diffraction grating? What happens to this upper limit as the wavelength is increased? What happens to it as the number of lines per unit length is decreased?

14. Suppose that you want to make a diffraction grating for infrared radiation of wavelength 10^{-4} meter. What grating spacing would be appropriate? What would be the appropriate spacing for microwaves with a wavelength of 10 cm?
15. Explain what is meant by Fresnel diffraction and contrast it with Fraunhofer diffraction.
16. A person looks at light through a very narrow space between two of his fingers. Does he see a Fresnel or a Fraunhofer diffraction pattern? Explain.
17. What happens to the maximum intensity of a Fraunhofer diffraction pattern as:
    (a) The wavelength is increased?
    (b) The slit width is increased?
18. Describe the nature of the single-slit Fraunhofer diffraction pattern that is obtained by use of a slit of width b equal to the wavelength of the light used.
19. What property of the lenses in telescopes is crucial in enabling us to distinguish between two very close stars? Explain why increasing the magnification of a telescope alone will not improve our ability to resolve two very close celestial objects.
20. It was discovered by Davisson and Germer that electrons of mass m and traveling at a velocity v will, under suitable conditions, behave as a *matter wave* with a wavelength

$$\lambda = \frac{h}{mv}$$

where h is Planck's constant ($= 6.63 \times 10^{-34}$ J-s). Explain, in qualitative terms, why an electron microscope—that is, a microscope that uses a stream of electrons instead of light rays—might enable us to see structures that cannot be resolved by ordinary microscopes.

21. The light from two very close celestial objects is viewed by use of a blue filter. Would the resolution be better or worse if a red filter were used instead?
22. Light from a diffuse source is incident on a Newton's rings apparatus as in Figure 32-23. Assuming that the plates A and B in the figure are both made of glass, compare the rings seen by reflection by the observer at O with those that are observed by transmission by an observer who looks upward from below plate A.
23. Suppose that a thin film of air, as in Figure 32-22, is exposed to an extended source of coherent light such as that supplied by a laser. In what way, if any, will the observed interference pattern differ from that obtained by using a monochromatic, but incoherent, extended source?
24. A soap film on a loop of wire is held so that the normal to the plane of the loop is horizontal. On being exposed to an extended light source, it is found that when viewed by reflected light, the top of the film is dark and various horizontal colored fringes appear below. Verify this experiment by use of materials directly available to you and explain the phenomenon.

PROBLEMS

1. What is the electric field amplitude E_0 associated with a monochromatic electromagnetic wave with an average energy flux of 2.0×10^{-2} W/m^2?
2. Two point sources are a distance 2.0×10^{-7} meter apart and radiate coherently and in phase at $\lambda =$ 5500 Å. With respect to an observer along the line joining them:
   (a) What is the phase difference between them?
   (b) What is the path difference between them?
3. Repeat both parts of Problem 2, but assume that the sources have a phase difference $\alpha_1 - \alpha_2 = 90°$. See (32-4).
4. Two sources S_1 and S_2 are separated by a distance Δ and located in the x-y plane at the points $(\pm\Delta/2, 0)$. Assuming that the sources radiate coherently and in phase at the wavelength λ, show that the condition for a maximum in intensity at the point (x, y) is

$$\sqrt{\left(x+\frac{\Delta}{2}\right)^2+y^2}-\sqrt{\left(x-\frac{\Delta}{2}\right)^2+y^2} = m\lambda$$

   where m is an integer. Prove that this curve is a hyperbola.
5. If, in the double-slit arrangement in Figure 32-9, the fringes that appear on the screen are a distance of 0.5° apart, find the distance d between the slits if the source emits the bright line in the spectrum of sodium with a wavelength of 5892 Å.
6. Suppose that in Problem 5 the interfering light beams are brought to a focus on a screen by a lens of focal length 50 cm. What is the distance on the screen between neighboring maxima in the center of the pattern?
7. Blue light, of wavelength 4700 Å, is used to illuminate two slits separated by a distance of 0.3 mm. How far from the slits must a screen be placed so that the fringes are 1 mm apart?
8. A double slit is illuminated by light of wavelength 6000 Å and it is found that the fringes on a screen 90 cm away are separated by a distance of 5.0 mm. A second source is now used with the same equipment and it is found that for it the fringes are separated by 5.7 mm. (a) What is the separation distance between the slits? (b) What is the wavelength of the second source?

*9. The radiation from two very narrow line sources S_1 and S_2 is projected onto a screen by a thin lens of focal length f; see Figure 32-26. Let I_1 be the average intensity on the screen at P when the radiation from S_2 is blocked off, and I_2 the corresponding intensity if S_1 is blocked.

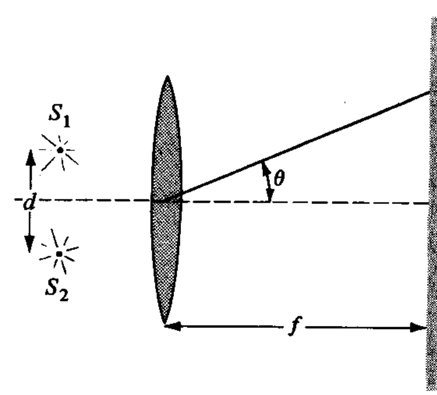


Figure 32-26

   (a) Show that if S_1 and S_2 are coherent, then when both are

radiating in phase the observed intensity $I(\theta)$ at P is

$$I(\theta) = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\left[\frac{2\pi d \sin\theta}{\lambda}\right]$$

(b) Show that for $I_1 = I_2 = I_0$ this agrees with (32-11).

10. Suppose the ratio I_1/I_2 of the intensities of the sources in Figure 32-26 is 2.5.
    (a) What is the ratio of the maximum to the minimum intensity observed on the screen?
    (b) If $\lambda = 4500$ Å and $d = 0.01$ mm, what is the angular separation between the maxima?
    (c) If $f = 40$ cm for the situation in (b), what is the spacing between maxima on the screen?

*11. Suppose that in a double-slit apparatus a very thin, transparent film of thickness t and refractive index n is placed in front of one of the slits. Show that the path difference Δ between the two beams is now given by

$$\Delta = d \sin\theta + t(n-1)$$

and write down the conditions for maxima and minima in this case.

12. In a two-slit experiment fringes are observed on a screen by use of light of wavelength 6500 Å. If a very thin, transparent film of refractive index 1.45 is placed in front of one of the slits, it is observed that the fringes are displaced upward by a distance of about three fringe widths; that is, the central fringe is now located at the point where formerly the maximum associated with a path difference of 3λ was located.
    (a) By how much has the film increased the optical path?
    (b) Making use of the result of Problem 11, find the thickness of the film.

13. Using the facts that

$$e^{\pm iy} = \cos y \pm i \sin y$$

and

$$1 + r + r^2 + \cdots + r^{N-1} = \frac{1-r^N}{1-r}$$

derive (32-8).

14. Show that for small values of θ the distance x_m of the mth maximum from the point P_0 on the screen in Figure 32-10 is

$$x_m = \frac{m\lambda D}{d}$$

and use this to derive (32-14).

15. By making use of the techniques of the calculus prove the following limits (N and m are integers):

(a) $$\lim_{x\to 0}\frac{\sin \pi x}{\pi x} = 1$$

(b) $$\lim_{x\to 0}\frac{\sin N\pi x}{\sin \pi x} = N$$

(c) $$\lim_{x\to m}\frac{\sin N\pi x}{\sin \pi x} = (-1)^{m(N-1)}N$$

*16. Show that the distance d between the two virtual images in the Fresnel biprism in Figure 32-12b is

$$d = 2a\alpha(n-1)$$

where n is the refractive index and where a and α are defined in the figure, with α a very small angle. Calculate the distance between neighboring maxima on a screen 100 cm from the biprism for the choices $n = 1.5$, $a = 15$ cm, $\alpha = 0.01$ rad, and $\lambda = 5000$ Å.

*17. Consider, in Figure 32-27, a plane wave incident at an angle i with respect to the normal onto a barrier, out of which have been cut two thin slits separated by a distance d. Show that the intensity $I(\theta)$ at the point P on the screen described by the angle θ is still given by (32-11),

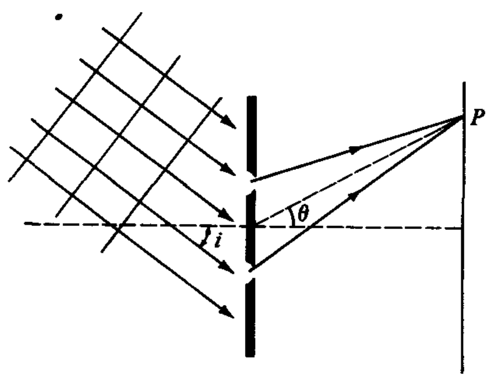


Figure 32-27

but with the replacement

$$d \sin\theta \to d(\sin\theta + \sin i)$$

18. Show that if monochromatic light of wavelength λ is incident at an angle i (with respect to the normal) on a diffraction grating, then the principal maxima occur at those values of θ given by

$$d(\sin i + \sin\theta) = m\lambda \qquad m = 0, 1, 2, \ldots$$

(*Hint*: Use the method suggested in Problem 17.)

19. A diffraction grating has 2000 lines/cm and is illuminated with red light of wavelength 6600 Å.
(a) What is the grating spacing?
(b) At what angles do the first two principal maxima occur?
(c) How many orders are observed?

20. A certain grating has 5000 lines/cm. How many orders are produced at a wavelength of (a) 4000 Å? (b) 5000 Å? (c) 7000 Å?

21. Yellow light from a sodium lamp of wavelength 5890 Å is incident normally on a diffraction grating, and five orders are observed. Assuming that the fifth maximum occurs at 88°, calculate the grating spacing.

22. Light of the wavelengths 6000 Å and 6300 Å falls normally on a grating that has 4000 lines/cm, and the diffracted light is brought to a focus on a screen by a thin lens of focal length 100 cm.
(a) What is the angular separation between the two beams in the first order?
(b) What is the separation distance on the screen between these two maxima?
(c) Repeat (b) for the third-order maxima.

23. An intense line is observed at an angle of 50° with respect to the normal of a diffraction grating of 6000 lines/cm. What are the possibile wavelengths of this line?

24. Make a plot of the formula for I in (32-17) as a function of the variable

$$\gamma = \frac{\pi d \sin\theta}{\lambda}$$

for the following cases: (a) $N = 2$; (b) $N = 3$; and (c) $N = 100$.

***25.** Consider the intensity formula for a diffraction grating in (32-17) as a function of the variable γ defined by

$$\gamma = \frac{\pi d \sin\theta}{\lambda}$$

(a) By using the results of Problem 15 show that the principal maxima are at the points

$$\gamma = m\pi \qquad (m = 0, 1, 2, \ldots)$$

and that the value of I here is

$$I = I_0 N^2$$

(b) Show that the zeros of I between the central and the first principal maximum occur at the points

$$\gamma = \frac{\pi}{N}, \frac{2\pi}{N}, \frac{3\pi}{N}, \ldots, \frac{(N-1)\pi}{N}$$

(c) Show that the secondary maxima are given by the roots of the equation

$$N \tan\gamma = \tan N\gamma$$

26. Light of wavelength 5400 Å is incident normally on a long slit of width 0.8 mm. If a lens of focal length 90 cm is placed behind the slit and a screen is placed in its focal plane:
(a) What is the distance from the center of the principal maximum to the first minimum?
(b) How far is it from the first minimum to the second minimum?

27. (a) Show by differentiation that the *maximum for the intensity* $I(\theta)$ in (32-20) occurs at the values of β satisfying

$$\tan\beta = \beta$$

(b) Show that if β_i is the value for β at the ith maximum, then the intensity I_i at this maximum is

$$I_i = I_m \cos^2\beta_i$$

28. For the physical situation in Problem 26, at what angle θ has the intensity dropped to half its value at the central maximum?

29. Make a plot, on the same graph, of the intensity distribution $I(\theta)$ as a function of θ produced by the Fraunhofer diffraction from a slit of width b if: (a) $\lambda = b$; (b) $\lambda = 0.1b$; and (c) $\lambda = 0.01b$. Interpret your results physically.

30. What is the angular diameter, as measured by θ_0, of the central maximum for each of the slit widths in Problem 29?

31. Show that if light is incident on a slit at an angle i with respect to its normal, then (32-20) for $I(\theta)$ is still applicable, except that the parameter β is now given by

$$\beta = \frac{\pi b}{\lambda}(\sin\theta + \sin i)$$

32. Light of wavelength 6200 Å is incident normally on a slit and projected onto a screen by a lens of focal length 75 cm. If the distance between the first two minima is 2.3 mm:
(a) How wide is the slit?
(b) How far away from the central maximum has the intensity dropped to half its central value?

33. Figure 32-28 shows the production of Fraunhofer diffraction due to plane waves of wavelength λ incident normally on two slits, each of width b and separated by a distance d. The intensity $I(\theta)$ at a point on the screen characterized by the angle θ is

$$I(\theta) = I_m \cos^2\gamma \frac{\sin^2\beta}{\beta^2}$$

where

$$\gamma = \frac{\pi d\sin\theta}{\lambda} \qquad \beta = \frac{\pi b\sin\theta}{\lambda}$$

(a) Show that for $b \ll \lambda$ this reduces to our previous results obtained for Young's experiment.

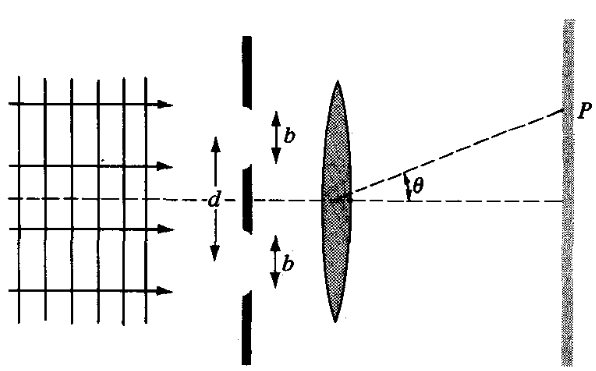


Figure 32-28

(b) For the special case $d \ll \lambda$, what should $I(\theta)$ reduce to? Show that it does.

(c) Show that for $d = b$, $I(\theta)$ reduces to the expected formula for Fraunhofer diffraction from a single slit of width $2b$.

34. For the graph in Figure 32-19c show that the minimum between the two peaks is approximately equal to $0.8I_m$, where I_m is the maximum intensity at a peak.

35. What is the minimum distance apart that two objects on the surface of the moon can be so that they can still be seen by the unaided eye? Assume that the pupil of the eye has a diameter of 4 mm, that the effective wavelength is 5500 Å, and that the distance to the moon is 3.8×10^5 km.

36. At what distance from an automobile can the unaided eye resolve its two headlights, assuming that they are 1.0 meter apart? Assume the data of Problem 35.

37. Two sunspots are a distance 8000 km apart. Assuming that the distance to the sun is 1.5×10^8 km and that the effective wavelength of light is 5500 Å, what is the diameter of the objective lens of a telescope that will just resolve the two spots?

38. What is the minimum angular separation between two distant stars that can just be resolved by the 200-inch telescope at Mount Palomar? Assume that the effective wavelength is 5500 Å.

39. A thin film of transparent material has a refractive index of 1.5 and a thickness of 10^{-4} cm. If it is illuminated with an extended source, what are the wavelengths of the fringes that appear in the reflected light? Assume normal incidence.

40. Suppose that in the apparatus in Figure 32-22 the glass plates are 10 cm long and the spacer is of such a thickness that the angle α between the plates is 2.0×10^{-4} rad. If the light is of wavelength 5000 Å and is viewed normally, calculate:

(a) The number of fringes that appear across the plate.

(b) The separation distance between the first two fringes.

41. Repeat Problem 40, but suppose that this time the space between the plates contains water of refractive index 1.33.

42. Suppose that in Figure 32-22 the upper plate has a refractive index of 1.4 and the lower one a refractive index of 1.6. Will there be a bright or a dark fringe along the line of contact between the two plates if: **(a)** The space between the plates contains air? **(b)** The space between the plates contains a material of refractive index 1.5? Justify your answer in each case.

43. A thin layer of oil of refractive index 1.4 lies on a glass plate of refractive index 1.5. What is the minimum thickness of the layer if it reflects very strongly light of wavelength 5000 Å incident from above? If a layer of oil of the same thickness lies between two glass plates, what wavelength will be strongly reflected? Assume normal incidence.

44. Generalize the results in (32-26) to the case where the angle of incidence θ is not zero (see Figure 32-21). Show in particular that if the film has a refractive index n and is in air, then for a maximum

$$2d\sqrt{n^2 - \sin^2\theta} = \left(m + \frac{1}{2}\right)\lambda$$

45. In a Newton's rings experiment, the radius of the fifth ring is found to be 0.2 cm for light of wavelength 6500 Å. **(a)** What is the radius of curvature of the lens? **(b)** What is the radius of the twentieth ring?

46. If the space between the plates in a Newton's rings experiment is filled with a material of refractive index n_0, show that the radii of the rings decrease by the factor $n_0^{-1/2}$.

47. An experiment with Newton's rings is carried out by use of two plano-convex lenses (see Figure 32-29). If the radii of curvature of the two lenses are R_1 and R_2, respectively, show that the radii of the maxima are given by

$$r_m = \left\{\lambda\left(m + \frac{1}{2}\right)\left(\frac{R_1 R_2}{R_1 + R_2}\right)\right\}^{1/2} \qquad m = 0, 1, 2, \ldots$$

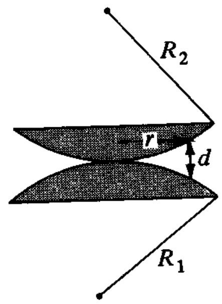


Figure 32-29

33 Polarization

Colors are the deeds and suffering of light.

GOETHE

33-1 Introduction

The phenomena of diffraction and interference are very generally associated with both longitudinal waves as well as transverse waves. As we saw in Chapter 18, for a longitudinal wave, such as a sound wave, the oscillations of the constituents of the medium are along the direction of propagation of the wave so that, as a rule, these waves are symmetric about this direction. By contrast, for a transverse wave the underlying vibrations are at right angles to the direction of propagation. Here, as exemplified by the traveling wave on the string in Figure 18-6, there may be an asymmetry about the direction of propagation of the wave. The term *polarization* is used to characterize a wave of this type, which is asymmetric about its direction of travel; the wave itself is said to be *polarized*. Polarization is a phenomenon associated exclusively with transverse waves; longitudinal waves do not exhibit polarization.

In Chapter 29 we saw that the electric and the magnetic vectors **E** and **B** of an electromagnetic wave are invariably perpendicular to the direction of travel of the wave. It follows that electromagnetic waves in general, and

light waves in particular, are transverse and should in principle exhibit polarization effects. This is indeed the case. The purpose of this chapter is to confirm this characteristic behavior of light by describing a number of methods that can be used to produce and to detect polarized light. Specifically we shall consider in detail the production of polarized light by (1) reflection and refraction; (2) selective absorption; and (3) double refraction or birefringence. This listing is by no means exhaustive; polarized light can be produced in other ways as well.

33-2 Linearly polarized light

As a preliminary to a discussion of practical methods for producing and observing polarized light, in this section and the next we shall describe the various states of polarization that are observed for light.

Consider, in Figure 33-1, an observer O viewing the radiation emitted by an accelerated charged particle. Assume for simplicity that the observer is far enough away from the particle so that the approaching wavefronts are planes and also that the acceleration $\mathbf{a}$ of the particle is perpendicular to the observation direction. According to (29-30), the electric vector $\mathbf{E}$ in the radiation that reaches the observer is proportional to, and oriented along, the direction of the acceleration $\mathbf{a}$. Hence, as shown, if $\mathbf{a}$ is along the vertical, then the direction of the electric field vector in the approaching wavefronts will also lie along this direction. The orientation of the associated $\mathbf{B}$-field is directed as shown and may be determined by the condition that the Poynting vector, $\mathbf{E} \times \mathbf{B}/\mu_0$, points along the direction of propagation of the wave. Furthermore, if the particle oscillates up and down at a certain frequency, then the electric vector in the observed radiation will similarly oscillate up and down with the same frequency. However, and this is an important point, the direction—although not the sense—of the electric field is unchanging and will in this case lie along the vertical throughout the particle's oscillations. Thus there is an asymmetry about the direction of propagation of the wave; in other words, the wave is *polarized.*

Any electromagnetic wave, such as this one, for which the direction, although not necessarily the sense, of the electric field is the same all along

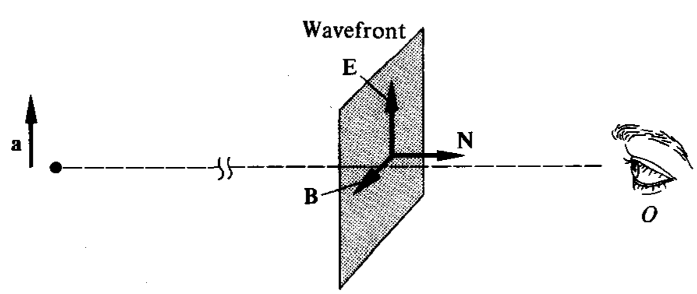


Figure 33-1

the wave is said to be *plane-polarized* or *linearly polarized* along that direction. For the situation in Figure 33-1, for example, this means that the light is linearly polarized along the direction of acceleration of the particle. Since the direction of the electric field **E** and that along which the wave propagates determine a unique direction for the associated **B**-field, in describing polarized light we shall from now on omit all reference to the latter field. Figure 33-2a shows the electric field vectors in the wavefront of an approaching wave that is plane-polarized along the vertical direction. The arrowheads at both ends of the lines represent the **E**-field and emphasize the fact that the sense of this field has no significance in this context. Figure 33-2b shows the corresponding case of light that is plane-polarized along the horizontal direction.

Figure 33-3 shows pictorially several light rays traveling to the right. The dots at various points along the ray in part (a) signify that the electric field vectors for this ray are perpendicular to the plane of the drawing. Hence this ray represents light linearly polarized perpendicular to the plane of the page. Correspondingly, the ray shown in Figure 33-3b represents light linearly polarized in the plane of the drawing. Since the electric field must be perpendicular to the light ray itself, the direction of polarization is along the vertical direction. Figure 33-3c shows how to represent the light resulting from the superposition of these two polarized light rays.

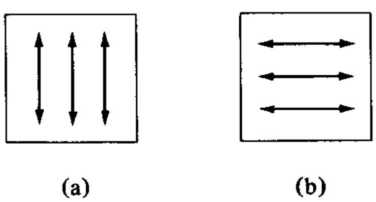


Figure 33-2

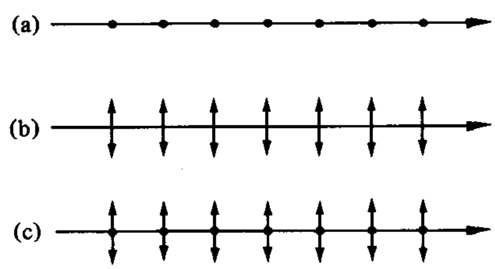
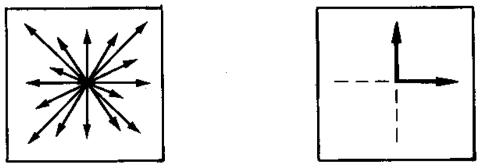


Figure 33-3

Let us consider now the radiation emitted by a collection of charged particles that separately accelerate in various directions. The observer, this time, will see radiation emitted by the totality of these particles and an instantaneous picture of the electric field vectors in a wavefront will now appear as in Figure 33-4a. In this case we say that the light is *unpolarized.* The distinction between unpolarized light and linearly polarized light should be carefully noted. For polarized light the *direction* of the electric vector in the wavefronts is constant in time at any fixed observation point, whereas for unpolarized light the direction of the electric vector changes, in a more-or-less random way, in the course of time.

There is an important relation between unpolarized and linearly polarized light, which may be established in the following way. Consider the instantaneous electric fields in the wavefront in Figure 33-4a and imagine taking their components along two mutually perpendicular directions. On adding



(a) (b)

Figure 33-4

these together, we obtain, as shown in Figure 33-4b, the components of the resultant vector along these two directions. Since this process may be repeated at any subsequent time, it follows that the originally unpolarized beam can be thought of as the superposition of two overlapping beams, each linearly polarized along one of the two mutually orthogonal directions. Note that this resolution of unpolarized light into two uncorrelated and linearly polarized components may be carried out along *any* two mutually perpendicular directions in the wavefront. It is for this reason that unpolarized light rays are represented, as in Figure 33-3c, by a ray with electric vectors oriented along two mutually perpendicular directions.

33-3 Circular and elliptical polarization

In order to describe states of polarization other than the above, a more quantitative characterization of unpolarized and polarized light is necessary. To this end, suppose that a monochromatic electromagnetic wave of wavelength λ is propagating along the z-axis of a certain coordinate system. As for the unpolarized wave in Figure 33-4, let us resolve the associated electric field in the wave into a vertical field $\mathbf{E}_1$ along the x-axis and a horizontal field $\mathbf{E}_2$ along the y-axis (see Figure 33-5). The total field is $(\mathbf{E}_1 + \mathbf{E}_2)$, where according to (29-20) through (29-23)

$$\begin{aligned} \mathbf{E}_1 &= \mathbf{i}E_0^{(1)} \sin\left[\frac{2\pi}{\lambda}(z - ct)\right] \\ \mathbf{E}_2 &= \mathbf{j}E_0^{(2)} \sin\left[\frac{2\pi}{\lambda}(z - ct) + \alpha\right] \end{aligned} \tag{33-1}$$

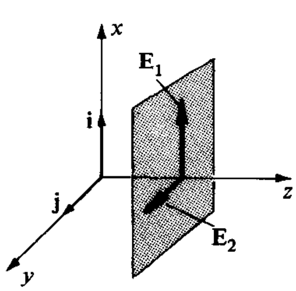


Figure 33-5

Here, **i** and **j** are unit vectors along the x- and y-axes, respectively, $E_0^{(1)}$ and $E_0^{(2)}$ are the amplitudes of the two component waves and α is the relative phase between them. The totality of all possible states of polarization of a light wave can be described completely in terms of these amplitudes $E_0^{(1)}$ and $E_0^{(2)}$ and the relative phase α.

If the two component waves in (33-1) are independent of each other, then the phase difference α between them must be *random*. In this case, the amplitudes $E_0^{(1)}$ and $E_0^{(2)}$ are generally equal to each other, and we say that the total wave described by $(\mathbf{E}_1+\mathbf{E}_2)$ is *unpolarized*. If α is random but $E_0^{(1)} \neq E_0^{(2)}$, then the wave is said to be *partially linearly polarized*. Generally speaking, unless special pains are taken, the light emitted by a source will be unpolarized, since by symmetry $E_0^{(1)}$ and $E_0^{(2)}$ must be equal.

If the phase difference α between the electric fields in (33-1) vanishes, then the associated wave is plane-polarized. For in this case the total field $(\mathbf{E}_1+\mathbf{E}_2)$ is

$$\mathbf{E}_1+\mathbf{E}_2=[\mathbf{i}E_0^{(1)}+\mathbf{j}E_0^{(2)}]\sin\left[\frac{2\pi}{\lambda}(z-ct)\right]$$

and since the direction of this field is constant in both space and time, it represents light that is linearly polarized along a certain direction. The angle θ between this polarization direction and the x-axis is easily found to be

$$\tan\theta=\frac{E_0^{(2)}}{E_0^{(1)}} \tag{33-2}$$

Besides unpolarized and linearly polarized light, (33-1) can also be used to characterize *circularly* and *ellipically* polarized light. If, for example, $E_0^{(1)}=E_0^{(2)}=E_0$ and $\alpha=-\pi/2$, then the sum of the fields in (33-1) is

$$\mathbf{E}=\mathbf{E}_1+\mathbf{E}_2=E_0\left\{\mathbf{i}\sin\left[\frac{2\pi}{\lambda}(z-ct)\right]-\mathbf{j}\cos\left[\frac{2\pi}{\lambda}(z-ct)\right]\right\} \tag{33-3}$$

since $\sin(\theta-\pi/2)=-\cos\theta$. To see the significance of this field, suppose that at a fixed instant t_0, the phase $2\pi(z-ct_0)/\lambda$ vanishes. It follows by substitution that at $t=t_0$, $\mathbf{E}=-\mathbf{j}E_0$. Similarly, a quarter of a period or $\lambda/4c$ later, the electric field will have the value $-\mathbf{i}E_0$ and at time $(t_0+\lambda/2c)$ its value will be $+\mathbf{j}E_0$. Hence, as shown in Figure 33-6a, at a fixed point z, during the course of time this electric vector rotates clockwise in the x-y plane. Its tip therefore traces out in a period λ/c a circle of radius E_0 with uniform angular velocity. The radiation associated with this field is said to be *right circularly* polarized. Correspondingly, *left circularly polarized* light results if the tip of the electric vector traces out a circle in the opposite sense. The choice $E_0^{(1)}=E_0^{(2)}=E_0$; $\alpha=+\pi/2$ in (33-1) describes this type of polarized light. The term *circularly polarized* light, without qualification, is used very generally to describe either type. Note that in these definitions the direction—clockwise or counterclockwise—in which the tip of the electric vector travels is with reference to the observer toward whom the wave is advancing.

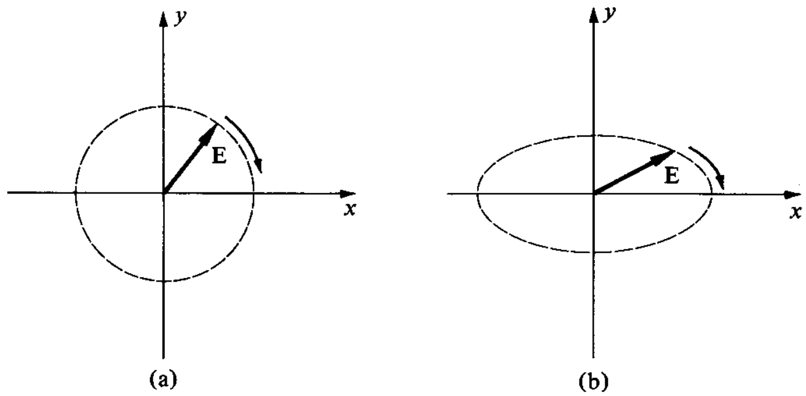


Figure 33-6

If the tip of the electric vector traces out an ellipse, then the associated light ray is said to be *elliptically polarized*; see Figure 33-6b. It is left as an exercise to confirm by use of arguments similar to those above that elliptically polarized light results if (1) $\alpha = \pm\pi/2$, but $E_0^{(1)} \neq E_0^{(2)}$; or (2) the phase angle α is not an integral multiple (including zero) of $\pi/2$. For example, the analogue of (33-3) for the case $\alpha = -\pi/2$ and $E_0^{(1)} \neq E_0^{(2)}$ is

$$\mathbf{E} = \mathbf{i}E_0^{(1)} \sin\left[\frac{2\pi}{\lambda}(z - ct)\right] - \mathbf{j}E_0^{(2)} \cos\left[\frac{2\pi}{\lambda}(z - ct)\right]$$

so that now at successive quarter-periods the electric vector assumes the values $-\mathbf{j}E_0^{(2)}$, $-\mathbf{i}E_0^{(1)}$, $+\mathbf{j}E_0^{(2)}$, and $+\mathbf{i}E_0^{(1)}$, respectively. Hence, assuming that $E_0^{(1)} > E_0^{(2)}$, the electric vector traces out clockwise an ellipse of semimajor and semiminor axes $E_0^{(1)}$ and $E_0^{(2)}$, respectively.

33-4 Polarization of light by reflection

For most of the remainder of this chapter we shall be concerned with methods for producing and detecting the various states of polarization defined above. One of the easiest methods to describe involves reflecting an unpolarized light beam at the interface between two dielectric media. This is the subject of this section.

Consider, in Figure 33-7a, a beam of *unpolarized* light incident, at an angle θ_1, onto the interface between, say, air and a medium of refractive index n. Part of the incident light will be reflected at the same angle θ_1 and the remainder will enter the medium at an angle θ_2 given by Snell's law:

$$\sin\theta_1 = n \sin\theta_2 \tag{33-4}$$

Now, although the amplitudes of the electric vectors—parallel and perpendicular to the plane of incidence—in the unpolarized incident ray are equal, this will *not*, in general, be true for the reflected and the refracted rays. The precise form of these amplitudes may be obtained by use of Fresnel's

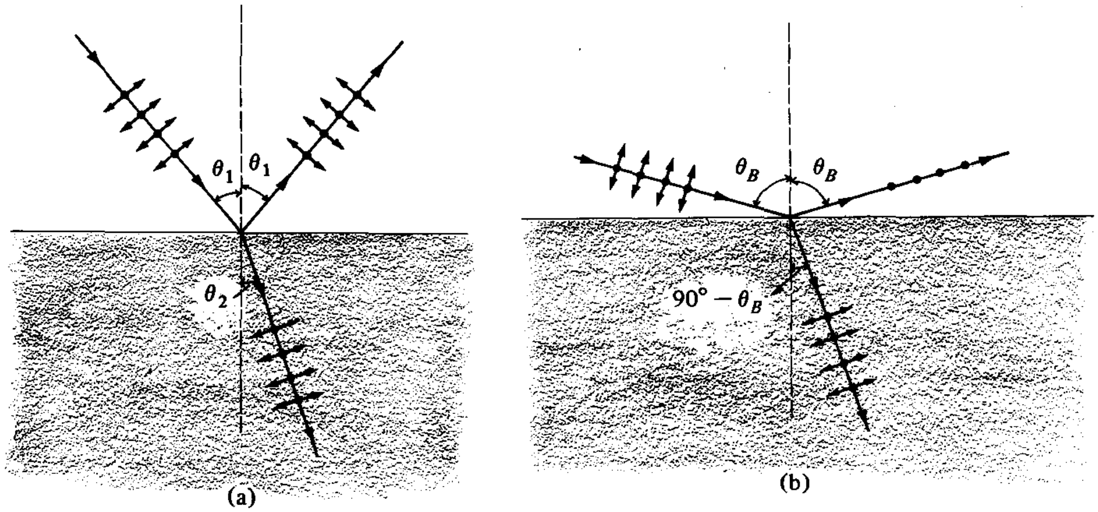


Figure 33-7

equations in (30-19) and (30-20). The first of these in (30-19) gives the amplitudes for the case of light polarized in the plane of incidence, and the second in (30-20) if the electric vector is perpendicular to the plane of incidence. It follows from these relations that the intensity of the reflected and the refracted light beams depends on the state of polarization of the incident ray as well as on the refractive index n and the angle of incidence θ_1. In general, for a given refractive index n the reflected and the refracted beams will be partially polarized, depending on the angle of incidence θ_1.

Of particular interest in this connection is a certain angle of incidence for which the reflected beam is totally linearly polarized along a direction perpendicular to the plane of incidence; see Figure 33-7b. The angle θ_B at which this occurs is known as ***Brewster's angle*** and may be thought of as that angle of incidence for which the angle between the reflected and the refracted rays is 90°. Reference to the first of Fresnel's equations in (30-19) shows that since tan 90° is infinite, at this angle θ_B, no light whose electric vector lies in the plane of incidence is reflected. Hence, if θ_1 has the value θ_B, so that $\theta_1 + \theta_2 = 90°$, then the electric vector in the reflected light must be perpendicular to the plane of incidence. Note that the transmitted beam need *not* be completely polarized even when the angle of incidence is θ_B. Reference to (30-20) shows that even at Brewster's angle, light which is plane-polarized perpendicular to the plane of incidence is partially reflected and partially transmitted. The above state of pure linear polarization of the reflected beam results only because there is no reflection of that part of the incident light which is polarized in the plane of incidence.

To obtain the relation between Brewster's angle θ_B and the refractive index n of the medium it is necessary to make use of Snell's law in (33-4). By definition of θ_B, the angle of refraction θ_2 associated with the angle of incidence $\theta_1 = \theta_B$ is

$$\theta_2 = 90° - \theta_B$$

Substitution into (33-4) leads to

$$\sin \theta_B = n \sin \theta_2 = n \sin (90° - \theta_B) = n \cos \theta_B$$

and thus

$$\tan \theta_B = n \qquad \textbf{(33-5)}$$

The corresponding formula when the incident ray travels in a medium of refractive index n_1 and strikes an interface with a medium of index n_2 is shown in the problems to be

$$\tan \theta_B = \frac{n_2}{n_1} \qquad \textbf{(33-6)}$$

Example 33-1 At what angle β above the horizon is the sun so that a person observing its rays reflected in water ($n = 1.33$) finds them linearly polarized along the horizontal (see Figure 33-8)?

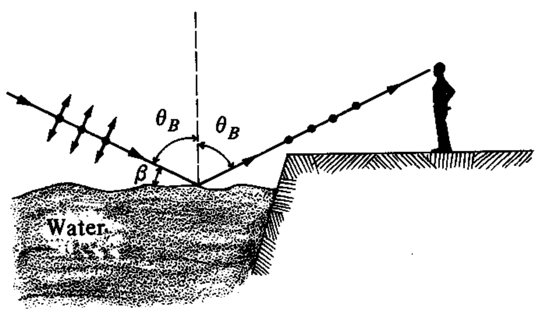


Figure 33-8

Solution In order for the reflected rays to be linearly polarized, the angle of incidence must be Brewster's angle. Substituting the value $n = 1.33$ into (33-5), we find that

$$\tan \theta_B = 1.33$$

and thus $\theta_B \cong 53°$. The sun's angle β above the horizon is the complement of this angle. Hence

$$\beta = 37°$$

Example 33-2 *Linearly polarized* light is incident at Brewster's angle on the surface of a medium. What can you say about the refracted and reflected beams if for the incident beam:

(a) The direction of polarization is parallel to the plane of incidence?
(b) The direction of polarization is perpendicular to the plane of incidence?

Solution

(a) At Brewster's angle the parallel component is completely refracted. Thus no light is reflected at all in this case.

(b) Here as may be verified by use of (30-20) some of the incident light is reflected and some is refracted. Both the reflected and the refracted beams will be polarized perpendicular to the plane of incidence just as was the incident beam. The direction of polarization of a light beam is, of course, not altered on being reflected or refracted.

33-5 Dichroic crystals

Suppose that a *linearly polarized* light beam is incident on the interface of a refractive medium at Brewster's angle. If the direction of polarization lies in the plane of incidence, then according to the results of Section 33-4, none of the incident light is reflected; it is entirely refracted into the second medium. Moreover, if the reflecting surface is rotated about the direction of the incident ray as an axis and in a way so that the angle of incidence remains θ_B throughout, a steady variation in the intensity of the reflected light is observed. This intensity vanishes when the direction of polarization lies in the plane of incidence and is a maximum when it is at right angles to this plane. The results of Section 33-4 therefore can be used not only to devise methods for producing linearly polarized light but also for detecting the existence of- and the state of polarization of such light rays.

In practice it is found much more convenient not to use these properties of reflected light, but rather to produce and to detect polarized light by other means. Generally speaking, these alternate and more practical methods involve the usage of certain anisotropic crystals whose optical properties depend on the orientation of the electric vector in the light beam being transmitted. There are two types of crystals that we shall consider in this connection. The first type is known as *birefringent* or *doubly refracting* crystals, examples of which are calcite and quartz. These crystals will be discussed in Section 33-7. The purpose of this and the next section is to describe a second type, known as *dichroic crystals.* An example of such a crystal is the mineral tourmaline.

A dichroic crystal is one in which there exists a certain direction in the crystal, known as the *axis of easy transmission,* which has the following property: If a light beam goes through the crystal, then the component of the beam polarized along this axis is transmitted without essential modification. By contrast, a light beam that is polarized perpendicular to this direction undergoes absorption inside the crystal. Note that the axis of easy transmission is a *direction* in the crystal and not an axis through it. Regardless of the path followed by a light beam in traversing a crystal, at any point of this path that component of the light polarized perpendicular to the line through this point and parallel to the axis of easy transmission will be absorbed.

Consider, in Figure 33-9, a beam of unpolarized light incident normally on the face of a dichroic crystal whose axis of easy transmission is parallel to that face. If the incident light is resolved into one polarization component along, and the other at right angles to, the axis of easy transmission, it follows that only the former is transmitted. The component of the beam with the electric field vector perpendicular to the axis of easy transmission is absorbed. Therefore, assuming that the crystal is thick enough so that this perpendicular component is completely absorbed, it follows that the light transmitted through this dichroic crystal will emerge as linearly polarized light with the electric field along the axis of easy transmission.

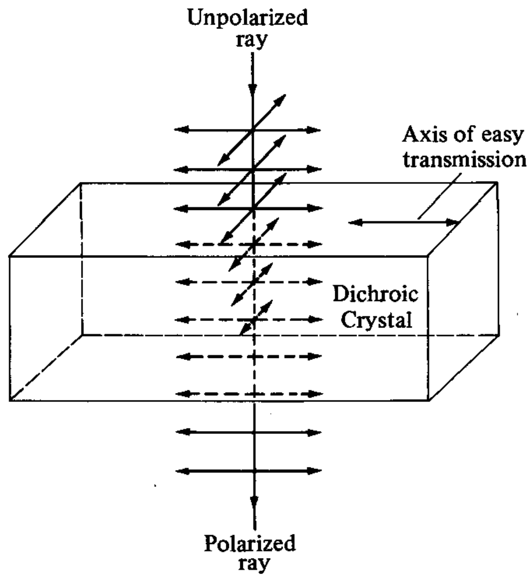


Figure 33-9

To illustrate the kind of physical processes involved in generating polarized light by use of dichroic crystals, let us consider briefly a certain substance developed by Edwin H. Land in 1938 and known as *Polaroid.* Polaroid normally comes in the form of thin plastic sheets and is made of a substance which contains long hydrocarbon chains to which are attached iodine atoms. In the manufacturing process, these thin sheets are stretched in a way so that these long molecular chains tend to line up in a parallel array along the direction of stretch. Associated with this regular structure are free electrons, which are not attached to any particular molecule and thus are free to travel, but only along the direction of alignment of the hydrocarbon chains. In particular they are *not* free to travel at right angles to this direction. If, now, as in Figure 33-10, a beam of unpolarized light is normally incident on such a sheet, these electrons can absorb energy only from that component of the radiation whose electric field is along the direction of the molecules. The electrons cannot move perpendicular to these chains, and hence cannot absorb the other component of the radiation. It follows that the component of the radiation polarized along the direction of the

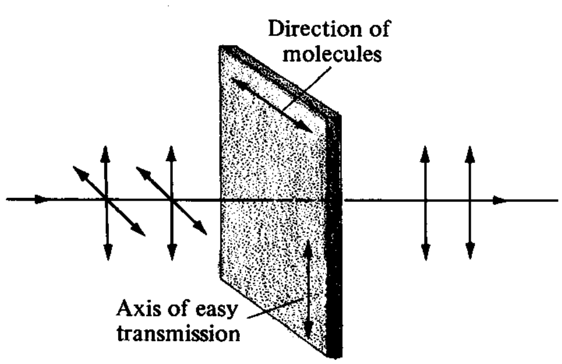


Figure 33-10

molecules will be absorbed and thus not transmitted. Therefore, as shown in Figure 33-10, the emerging light is linearly polarized in a direction perpendicular to that along which the molecules are aligned. The axis of easy transmission must therefore be perpendicular to this direction of the molecules.

33-6 Malus' law

Consider, in Figure 33-11, an observer O looking into a unpolarized light beam, which has been transmitted through two parallel Polaroid sheets with axes of easy transmission mutually perpendicular. The light transmitted through the first Polaroid will emerge linearly polarized along the vertical direction. Since the axis of the second Polaroid is horizontal, while the light incident on it is polarized along the vertical, it follows that the beam will be completely absorbed. Thus, the observer O will see no light! Experiment shows that he will actually see a small amount of blue light if it is present in the incident beam. This feature follows from the fact that Polaroid is not completely effective at all wavelengths.

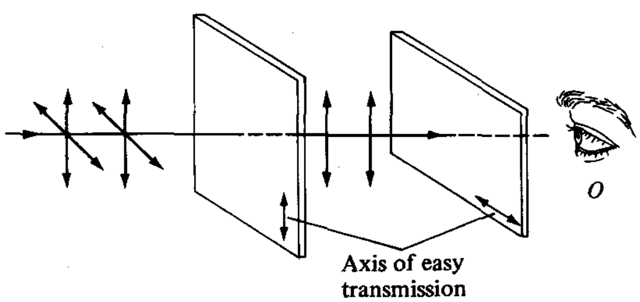


Figure 33-11

Figure 33-12 shows a similar experiment, but this time with the axes of the parallel sheets of Polaroid at an angle θ relative to each other. Let us assume again that unpolarized light is incident from the left, and that the first Polaroid sheet, which is known as the *polarizer* when used in this way, has its axis of easy transmission at the angle θ with respect to the vertical. If the

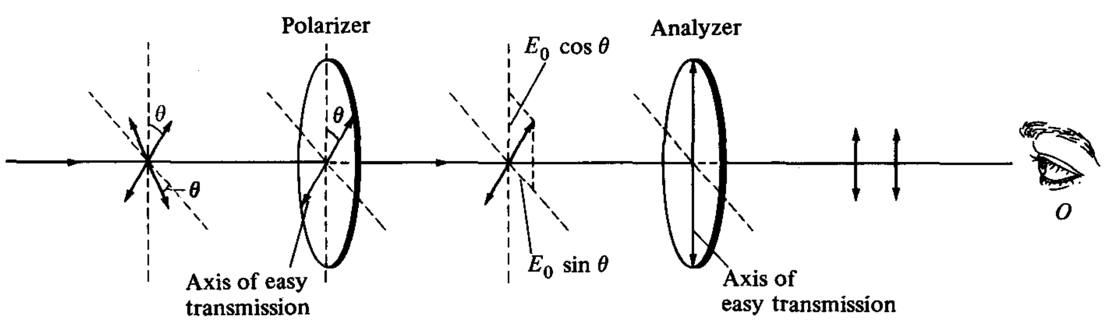


Figure 33-12

incident light is resolved into a component polarized along the axis of the polarizer, and a second at right angles to this axis, only the former will be transmitted. Hence the light emerging from the polarizer is polarized along a line that makes an angle θ with the vertical. Its intensity—neglecting any reflections or absorption—will be exactly half the incident intensity. The amplitude E_0 of the transmitted beam is the same as the amplitude of the corresponding component of the incident beam.

Let us now consider what happens to the beam on going through the second Polaroid sheet, or *analyzer.* The incident polarized light of amplitude E_0 may be thought of as the superposition of two polarized beams: one of amplitude $E_0 \cos\theta$ and polarized along the vertical; and the second of amplitude $E_0 \sin\theta$ polarized along the horizontal. Since the latter will be absorbed by the analyzer, it follows, as shown in the figure, that the transmitted beam will be polarized along the vertical and have the amplitude $E_0 \cos\theta$. According to (29-25) the intensity I of the transmitted beam is proportional to the square of this amplitude $E_0 \cos\theta$. For the same reason, the intensity I_0 of the radiation incident on the analyzer is proportional to $|E_0|^2$. Hence follows the basic relation

$$I = I_0 \cos^2\theta \tag{33-7}$$

which is known as *Malus' law.*

In deriving (33-7) we have explicitly assumed that the component of the incident beam that is polarized along the axis of easy transmission undergoes no reflection or absorption while going through the Polaroid. If we interpret the quantity I_0 in (33-7) not as the intensity of the radiation incident on the analyzer but rather as the maximum intensity observed when the axis of the polarizer and analyzer are parallel, then (33-7) is valid in general. For example, it is applicable if in place of two Polaroid sheets, we use two reflecting surfaces and reflect the light at Brewster's angle in each case. The term "Malus' law" is usually reserved for (33-7) when I_0 has this more general meaning.

Example 33-3 If a third Polaroid sheet with its axis at an angle θ with respect to the vertical is inserted between the two Polaroid sheets in Figure 33-11, what intensity will the observer O see now? Assume that the amplitude of the light transmitted through the first Polaroid is E_0 and that reflection and absorption losses are negligible.

Solution The light incident on the middle Polaroid sheet has the amplitude E_0 and is polarized along the vertical. According to the above arguments, the radiation transmitted through it will have the amplitude $E_0 \cos\theta$ and be polarized along the axis of this sheet, that is, along a direction making an angle θ with respect to the vertical. Repeating the above argument a second time, we find that the light emerging from the third sheet will have the amplitude $(E_0 \cos\theta) \sin\theta$ (since the axis of easy transmission of the third sheet is horizontal). Hence the observed intensity I is proportional to $[\cos\theta \sin\theta]^2 = (\sin^2 2\theta)/4$.

By contrast to this very striking result, whose validity is easily confirmed experimentally, if the added Polaroid sheet is placed in front of or behind the two Polaroid sheets in Figure 33-11 the observer will see no light for any value of the angle θ. Can you explain this physically?

Example 33-4 Suppose a beam of circularly polarized light is normally incident on a thin Polaroid sheet.

(a) What is the nature of the transmitted light?

(b) What happens to the observed intensity if the Polaroid sheet is rotated about the direction of the beam?

Solution

(a) Circularly polarized light, as we have seen, is the superposition of two light beams that are polarized along two orthogonal directions and have a phase difference of 90° between them. It follows that, just as for linearly polarized light, if such a beam goes through a Polaroid sheet, one of its two components will be absorbed and the other transmitted. Hence the emerging beam will be linearly polarized in a direction parallel to the axis of easy transmission of the sheet.

(b) Since the electric field vectors in two components of a circularly polarized beam have the same amplitudes and the phase difference of $\pm 90°$, it follows that the transmitted intensity is the same regardless of the orientation of the axis of easy transmission. Hence no variation in intensity would be observed as the Polaroid sheet is rotated. This is to be contrasted with the case of elliptically polarized light, for which there would be a variation in intensity as the sheet is rotated. Why?

33-7 Double refraction

The optical properties of a transparent, homogeneous, and isotropic medium, such as glass, are determined exclusively by its refractive index n. Thus a knowledge of n and its possible variation with wavelength makes possible—by use of Snell's law in (30-9) and Fresnel's equations in (30-19) and (30-20)—the calculation of the intensities and the directions of travel of light reflected and refracted at the surface of such a medium. Moreover, the velocity of light, c/n in such a medium, has the same value regardless of the direction of propagation. Of particular significance in this connection is the fact that this speed is the same for all states of polarization of the light beam.

To understand the physical basis for this independence of the speed of light in an isotropic medium to the direction of the electric field, let us recall (30-7):

$$n = \sqrt{\kappa} \tag{33-8}$$

which relates the refractive index of the medium to its dielectric constant. For a given electric field, as we saw in Chapter 22, the constituent dipoles of the medium tend to line up *along* the direction of the field. The extent to which this alignment takes place is measured by the strength of the dielectric constant κ in accordance with (22-29). Since, for a homogeneous and isotropic medium, κ is the same everywhere regardless of the direction or

the strength of **E**, it follows from (33-8) that the velocity of a light wave in such a medium is also independent of the electric field.

There is another class of materials of considerable interest, which are also macroscopically homogeneous, but are not optically isotropic. These anisotropic crystalline materials are called *birefringent* or *doubly refracting*. Mainly for reasons of simplicity, in the following we shall be concerned only with the particular class of these materials known as *uniaxial* crystals. Some examples of uniaxial crystals are calcite ($CaCO_3$) and quartz (SiO_2). The significance of the term *uniaxial* will be clarified below.

Consider a light wave traversing a birefringent crystal. Experiment shows that its speed depends, in general, on both the direction of propagation of the wave and on its state of polarization. Most striking is the fact that on entering the crystal, an initially unpolarized light beam will split into two linearly polarized and distinct beams. One of these is called the *ordinary ray* or *o-ray* and the other the *extraordinary ray* or *e-ray*. The behavior of the *o*-ray is essentially the same as that of a light wave in an isotropic medium. For example, on entering the birefringent crystal the *o*-ray is refracted in accordance with Snell's law. Also the refractive index n_o of the *o*-ray as well as its speed c/n_o in the crystal is the same regardless of the direction of travel of the *o*-ray. By contrast, the *e*-ray behaves in a most unusual way. Its velocity of propagation is different for different directions in the crystal! Moreover, its direction of travel, on being refracted into the crystal, is *not* generally consistent with Snell's law. It is very convenient nevertheless to associate a certain refractive index n_e with the *e*-ray. If $n_e > n_o$ the crystal is said to be *positive*, whereas if $n_e < n_o$ the crystal is said to be *negative*. As will be seen in Section 33-8, where a precise definition of n_e is given, for a positive crystal, such as quartz, c/n_e is the smallest value for the velocity of the *e*-ray in the crystal, while for a negative one, such as calcite, c/n_e is the maximum speed of the *e*-ray. The speed of the *o*-ray in a positive crystal always exceeds that of the *e*-ray, whereas for a negative crystal the speed of the *e*-ray is never less than that of the *o*-ray. Depending on its direction of travel, the velocity of the *e*-ray in a positive crystal varies from a maximum value of c/n_o to a minimum of c/n_e, and for a negative crystal the velocity ranges from a minimum of c/n_o to a maximum of c/n_e. Note that the velocity of the *o*-ray is the same in all directions in the crystal.

Table 33-1 lists values for n_o and n_e for a number of positive (p) and

Table 33-1 Indices of refraction of birefringent crystals at 5893 Å

Substance	n_o	n_e
(n) Calcite ($CaCO_3$)	1.658	1.486
(n) Sodium nitrate ($NaNO_3$)	1.587	1.336
(p) Zinc sulfide (ZnS)	2.356	2.378
(p) Quartz (SiO_2)	1.544	1.553
(p) Ice (H_2O)	1.309	1.313

negative (n) crystals. Since for ice the difference ($n_e - n_o$) is very small, care must be exercised to see double refraction by use of this substance.

Figure 33-13 shows an experiment that demonstrates double refraction. An unpolarized light beam is normally incident on one side of a doubly refracting medium with parallel opposite faces. At the point of entry, the beam splits into two rays: an o-ray and an e-ray. The former obeys the laws of geometrical optics and thus is transmitted without deviation. By contrast, the e-ray is bent away from the normal when entering the crystal and toward the normal on leaving. It thus emerges in a direction parallel to and displaced from the incident beam. Of considerable interest is the fact that the emerging rays are linearly polarized along mutually perpendicular directions. This feature, which will be examined in more detail in Section 33-8, is the reason that doubly refracting crystals are so useful in a study of polarized light.

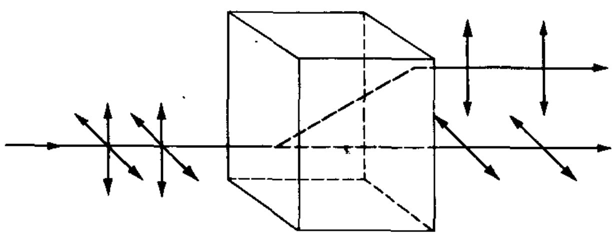


Figure 33-13

33-8 Polarization in birefringent crystals

To obtain a physical understanding of this distinctive behavior of doubly refracting crystals it is convenient to make use of the same physical ideas used to explain the optical behavior of isotropic materials. Since for isotropic substances, the constituent dipoles always tend to line up along the direction of the external field, the induced dipole moment per unit volume **P** is invariably parallel to **E**. By contrast, in a doubly refracting crystal, **P** does *not* generally lie along the direction of the electric field. Hence some modification of (22-29) and (22-30) is required in order to describe this unique behavior.

More detailed studies show that regardless of the complexity of a nonabsorbing crystal, there are always three mutually perpendicular directions in a crystal along which **P** will be parallel to the electric field. These three directions are known as the *principal axes of the crystal* and a distinct dielectric constant is associated with each one. Let us define a Cartesian coordinate system with axes along the principal axes of the crystal and let κ_1, κ_2, and κ_3 be the respective dielectric constants associated with the x-, y-, and z-axes; see Figure 33-14. If, for example, **E** is an electric field along the x-axis, the associated dipole moment **P** will be given by (22-29) as $\epsilon_0(\kappa_1 - 1)\mathbf{E}$ with dielectric constant κ_1. Similarly, if **E** is along the y-axis, the appropriate dielectric constant is κ_2. The significance of the fact that these dielectric

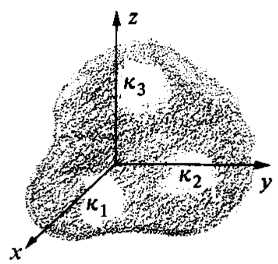


Figure 33-14

constants are different is that the respective indices of refraction n_1, n_2, and n_3 (in accordance with (33-8)) will also be different. Thus if a linearly polarized electromagnetic wave with its electric vector along the x-axis travels through the crystal, its speed will be $c/n_1 = c/\sqrt{\kappa_1}$. Similarly, if the wave is polarized along the y-axis or the z-axis, its speed will be $c/\sqrt{\kappa_2}$ or $c/\sqrt{\kappa_3}$, respectively. For polarization directions not along one of these three principal axes the velocity of the associated wave assumes an intermediate value.

The important feature to be noted here is that the *velocity of light in the crystal is determined by the direction of the electric field in the wave*; the propagation direction of the wave is related to the speed only indirectly through its relation to the direction of polarization. It is for this reason that on entering a birefringent crystal, an originally unpolarized beam will split into an o-ray and an e-ray, each characterized by its distinct state of polarization.

A crystal for which no two of the principal dielectric constants κ_1, κ_2, and κ_3 are equal is called a *biaxial crystal.* Feldspar, gypsum, and mica are examples of such crystals. The name *biaxial* has to do with the fact that for these crystals there are two distinct directions in the crystal along which the o-ray and the e-ray can travel at the same speed. These directions, which are known as the *optic axes* of the crystal, do *not* lie along any of the three principal axes.

In the following we shall confine our attention to a simpler class of crystals known as *uniaxial,* of which the substances listed in Table 33-1 are members. A uniaxial crystal is one for which two of the dielectric constants are the same. As the name implies, a uniaxial crystal has only one *optic axis* and this lies along that principal axis with the "unequal" dielectric constant. As for biaxial crystals, along the direction of the optic axis the speeds of the o-ray and the e-ray are precisely the same. Just as for the axis of easy transmission in dichroic crystals, the optic axis is a *direction* through the crystal and not an axis. Only along the unique *direction* of the optic axis is there no distinction between the o-ray and the e-ray.

Continuing our analysis of uniaxial crystals, let us assume to be specific that $\kappa_1 = \kappa_2 \equiv \kappa_0 \neq \kappa_3$ in Figure 33-14. The fact that the optic axis then lies along the z-axis in the crystal can be seen in the following way. The electric field associated with a wave traveling along the z-axis must lie in the x-y

plane. Hence, since $\kappa_1 = \kappa_2 = \kappa_0$, the velocity of this wave is $c/\sqrt{\kappa_0}$ and is independent of the state of polarization. We define the ordinary refractive index n_o of this crystal so that c/n_o is the velocity of a light wave traveling along the optic axis. According to (33-8), we have thus

$$n_o = \sqrt{\kappa_0} \tag{33-9}$$

Consider now an unpolarized electromagnetic wave traveling along the x-axis of this same crystal. This wave may be thought of as the superposition of two components: one linearly polarized along the y-axis and the other along the z-axis. The dielectric constant associated with the former is κ_0, so this component travels at the velocity c/n_o in accordance with (33-9). It is the ordinary ray, or o-ray. The direction of polarization of this o-ray is *perpendicular to the plane defined by the propagation direction and the optic axis.* By contrast, the other component, which is polarized along the z-axis, travels at the velocity corresponding to the dielectric constant κ_3. It is the extraordinary ray, or e-ray, and the associated refractive index n_e is

$$n_e = \sqrt{\kappa_3} \tag{33-10}$$

The direction of polarization of the e-ray is *in the plane determined by the optic axis and the direction of travel of the ray.* This latter feature, and the corresponding one stated above for the polarization direction of the o-ray, are very generally true; their validity is not restricted, as here, to light traveling along the principal axes of the crystal.

It is important to note that regardless of the direction of travel of an initially unpolarized light beam through a uniaxial crystal, there must be associated with it a component of the electric field along the x-axis or the y-axis. Hence any ray traveling through a uniaxial crystal will have an o-ray component whose velocity c/n_o is independent of the direction of travel. Unless the direction of the beam is along the optic axis itself, an e-ray will also be associated with it. However, the velocity of the e-ray varies with the direction in the crystal.

To determine the velocity of the e-ray along an arbitrary direction in the crystal it is necessary to carry out studies analogous to those described in Chapter 29. The results may be summarized in the following way. Imagine a pulse of light generated at some point O in a uniaxial crystal, and let u_x, u_y, and u_z be the components along the principal axes of the velocity $\mathbf{u}$ of one of these rays. Since the o-ray goes out with the same speed c/n_o in all directions, it follows that 1 second later the wavefront associated with the ordinary ray will be at the distance $|\mathbf{u}|$ from O, where

$$u_x^2 + u_y^2 + u_z^2 = \frac{c^2}{n_o^2} \qquad (o\text{-ray}) \tag{33-11}$$

If we think of a set of "velocity axes" with directions parallel to the principal axes of the crystal, (33-11) represents a sphere of radius c/n_o centered at the origin of this system. The velocity of an o-ray in any direction is the vector

along that direction from the origin to the point of intersection with the surface. Correspondingly, the surface associated with the e-ray is the ellipsoid of revolution

$$\frac{u_x^2}{c^2/n_e^2}+\frac{u_y^2}{c^2/n_e^2}+\frac{u_z^2}{c^2/n_o^2}=1 \qquad (e\text{-ray}) \tag{33-12}$$

and again the velocity of an e-ray in any direction is given by a vector along that direction from the origin to the point of its intersection with the ellipsoid. For an e-ray traveling along the z-axis, for example, we find by setting $u_x = u_y = 0$ in (33-12) that $u_z = c/n_0$, thus confirming that along the optic axis the e-ray and the o-ray travel at the same speed. Similarly, for an e-ray traveling along the x-axis, (33-12) predicts a speed c/n_e. Moreover, by use of (33-12) it is also possible to calculate the velocity of an e-ray in a direction other than a principal axis. The result, as shown in the problems, is that the speed u of an e-ray traveling along a direction making an angle θ with optic axis is

$$u=\frac{c}{[n_o^2\cos^2\theta+n_e^2\sin^2\theta]^{1/2}} \tag{33-13}$$

Figure 33-15 shows a cross section (of a plane containing the optic axis) through the surfaces defined by (33-11) and (33-12). In each case the optic axis is the line AB. Figure 33-15a shows the case of a positive crystal for which the e-ray ellipsoid is inside the o-ray sphere. Note the directions of polarization associated with a typical o-ray and e-ray. Figure 33-15b shows the corresponding situation for a negative crystal.

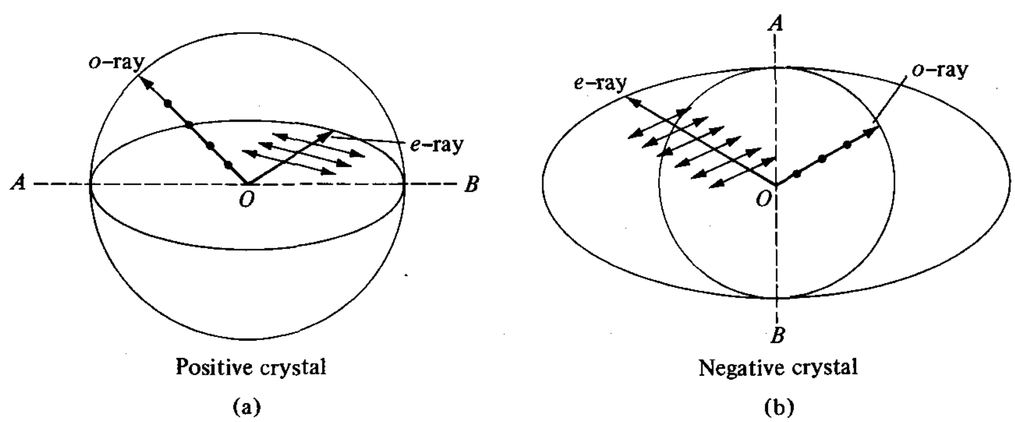


Figure 33-15

Figure 33-16 shows the formation of wavefronts in a negative uniaxial crystal as a result of unpolarized light being normally incident on a crystal face. Figures 33-16a and b consider the case when the optic axis is, respectively, parallel and perpendicular to the refracting surface. Figure 33-16c shows the corresponding case when the optic axis is perpendicular to the plane of the drawing. Finally, in Figure 33-16d we consider the more general case, where the optic axis makes an arbitrary angle with the surface.

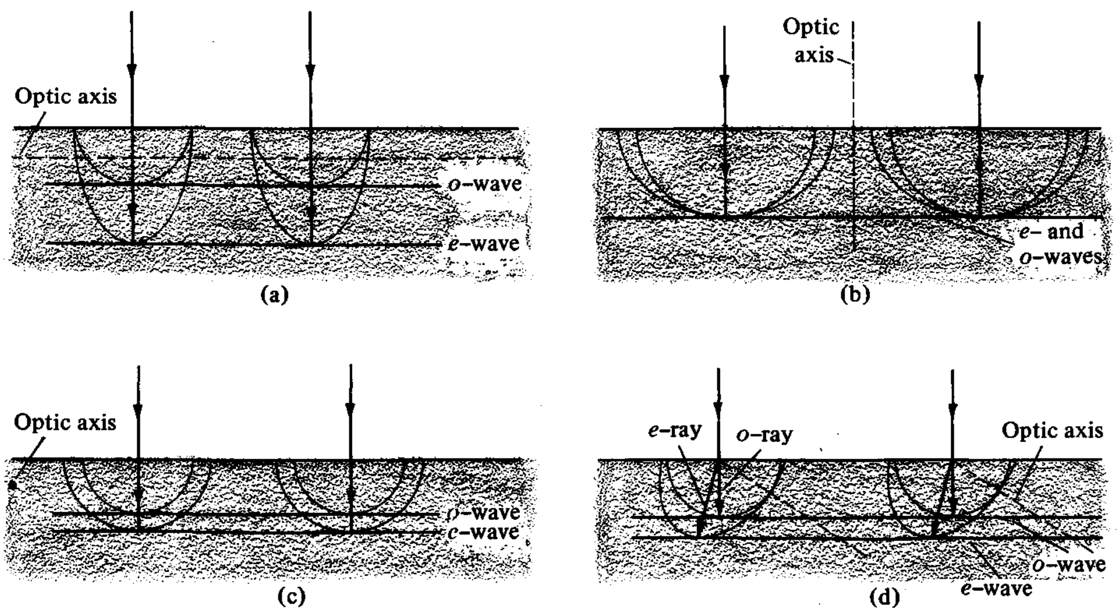


Figure 33-16

As shown by the arrows to the respective *e*- and *o*-wavefronts, on entering the crystal the two rays split up. This is the situation envisioned in Figure 33-13.

Example 33-5 Calculate the velocity in a calcite crystal of a light ray traveling at an angle of 45° with respect to the optic axis for:

(a) The *o*-ray.

(b) The *e*-ray.

Solution

(a) The velocity **u** of the *o*-ray is the same in all directions. Hence, by use of Table 33-1,

$$u = \frac{c}{n_o} = \frac{3.0 \times 10^8 \text{ m/s}}{1.658} = 1.8 \times 10^8 \text{ m/s}$$

(b) Because of the symmetry under rotations about the optic axis, we can assume that the *e*-ray lies in the x-z plane and makes equal angles with the x- and z-axes. Hence, setting $u_y = 0$ and $u_x = u_z$ in (33-12), we find that

$$u_x = \frac{c}{\sqrt{n_e^2 + n_o^2}} = \frac{3.0 \times 10^8 \text{ m/s}}{[(1.486)^2 + (1.658)^2]^{1/2}} = 1.35 \times 10^8 \text{ m/s}$$

Hence the speed u is

$$u = [u_x^2 + u_z^2]^{1/2} = \sqrt{2}u_x = 1.9 \times 10^8 \text{ m/s}$$

The same result can also be obtained by substitution into (33-13).

33-9 Quarter- and half-wave plates

One application of doubly refracting crystals is to the production and detection of circularly and elliptically polarized light. The purpose of this section is to describe a particular optical device known as a *quarter-wave plate,* which is very useful for this purpose.

Consider, in Figure 33-17, a light wave linearly polarized along a direction making an angle θ with the vertical and normally incident on a thin, doubly refracting sheet of thickness d and with the optic axis along the vertical. Let us decompose the incident electric field amplitude E_0 into a component $E_0 \cos\theta$ along the optic axis and a component $E_0 \sin\theta$ at right angles to it. The component along the optic axis is destined to become the amplitude of the e-ray and travels through the crystal at the speed c/n_e, and the other component is the amplitude of the o-ray which travels at the speed c/n_o. Because of this difference in velocity for the two rays, on emerging from the strip, there will be a certain phase difference δ between them. The emerging ray will in general no longer be plane-polarized as it was prior to entering the strip.

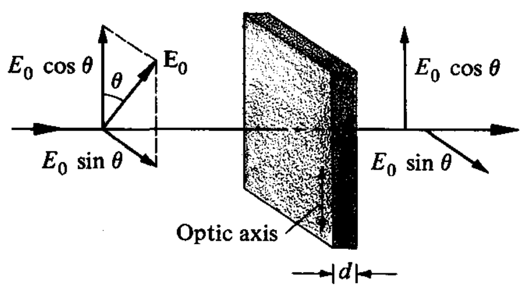


Figure 33-17

Now assuming that the incident wave is monochromatic with free-space wavelength λ, it follows from (29-24) that the wavelength of the o-ray in the medium is λ/n_o while that for the e-ray is λ/n_e. The phase difference δ that arises between these rays as a result of this wavelength difference may be obtained, by use of (32-7), as

$$\delta = \frac{2\pi}{\lambda/n_o} d - \frac{2\pi}{\lambda/n_e} d$$

so that

$$\delta = \frac{2\pi}{\lambda} d(n_o - n_e) \qquad \textbf{(33-14)}$$

Thus, assuming no absorption, the light ray that emerges from the biefringent plate in Figure 33-17 consists of two linearly polarized components, of the respective amplitudes $E_0 \cos\theta$ and $E_0 \sin\theta$, and with the phase difference δ. Comparing this with the definitions of linearly, circularly, and

elliptically polarized light in Section 33-3, we may draw the following conclusions:

1. If the thickness d of the plates is such that $\delta = \pm\pi/2$, and $\theta = 45°$ (so the amplitudes $E_0 \cos\theta$ and $E_0 \sin\theta$ are equal), then the emerging beam in Figure 33-17 is circularly polarized.
2. If the thickness d is such that $\delta = \pm\pi$, then the emerging beam is linearly polarized in a direction making an angle 2θ with respect to the incident direction; that is, the direction of polarization is rotated by an angle 2θ in this case.
3. If the thickness d is such that δ has a value other than an odd or even integral multiple of $\pi/2$, then the emerging beam is elliptically polarized.

The details of the arguments justifying these conclusions are found in the problems.

A *quarter-wave plate* is defined to be one for which the thickness d in (33-14) corresponds to a phase difference δ of $\pm 90°$. The importance of the quarter-wave plate is that it can be used as a tool to convert plane-polarized light to circularly polarized light, and conversely. If the direction between the optic axis of the quarter-wave plate and the direction of polarization of the incident radiation is 45°, the emerging ray will be circularly polarized. At other angles the emerging beam will be elliptically polarized.

A related optical device is the *half-wave plate.* It differs from a quarter-wave plate in that its thickness d is such that the phase difference δ between the emerging rays is 180°. The thickness of certain commercially available cellophane happens to be such that for certain wavelengths it is a reasonable approximation to a half-wave plate.

It should be noted that the phase difference δ between the emerging rays in Figure 33-17 depends on wavelength. Hence, in general, a given *quarter-wave plate* is useful only for an appropriately chosen color. For a given birefringent material, quarter-wave plates designed for different colors will have different thicknesses, according to (33-14).

Example 33-6 Describe the nature of the light transmitted through a quarter-wave plate if the incident light is circularly polarized.

Solution In accordance with (33-3), circularly polarized light is the superposition of two waves of equal amplitude, linearly polarized along perpendicular directions and with a phase difference of 90°. As the light passes through the quarter-wave plate, this phase difference becomes zero or 180° since the effect of the plate is to change one of the trigonometric functions in (33-3) to the other or to its negative. In either event, the emerging beam comes out linearly polarized along a direction making an angle of 45° with the optic axis. If viewed through a sheet of Polaroid, the emerging light will vary in intensity from zero to some maximum as the quarter-wave plate is rotated about the direction of the beam. This procedure is therefore a definite test for circularly polarized light.

Example 33-7 Consider, in Figure 33-18, an optical system consisting of a half-wave plate between two sheets of Polaroid whose axes of easy transmission are perpendicular. Assuming that the incident beam is unpolarized, what variation in intensity is observed by the observer O as the half-wave plate is rotated about the axis of the beam?

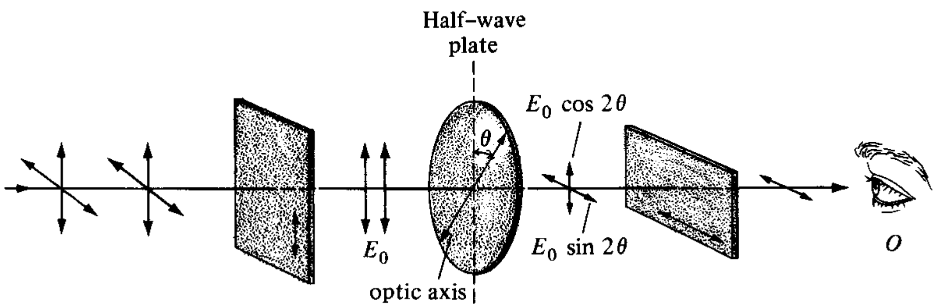


Figure 33-18

Solution The light ray emerging from the first Polaroid sheet is linearly polarized along the vertical. Thus if it were not for the *half-wave* plate, the observer at O would see nothing. However, as shown in the figure, the half-wave plate—assuming that its optic axis makes an angle θ with the vertical—has the effect of rotating the electric vector by an angle 2θ. Hence the beam emerging from the half-wave plate may be thought of as the superposition of two linearly polarized waves: one of amplitude $E_0 \cos 2\theta$ and polarized along the vertical, and the other at right angles with amplitude $E_0 \sin 2\theta$. The observer at O will see the horizontally polarized component with intensity proportional to $\sin^2 2\theta$.

As the half-wave plate is now rotated about the direction of the beam, the observed intensity will vary with θ as $\sin^2 2\theta$. It thus vanishes when the optic axis of the half-wave plate is parallel to the axis of easy transmission of the first Polaroid sheet, rises to a maximum when this angle is 45°, and vanishes when it is 90°, and so forth.

33-10 Summary of important formulas

If an electromagnetic wave of wavelength λ is traveling in free space, then the total electric field $(\mathbf{E}_1 + \mathbf{E}_2)$ in the advancing wavefront may be written

$$\begin{aligned} \mathbf{E}_1 &= \mathbf{i}E_0^{(1)} \sin\left[\frac{2\pi}{\lambda}(z-ct)\right] \\ \mathbf{E}_2 &= \mathbf{j}E_0^{(2)} \sin\left[\frac{2\pi}{\lambda}(z-ct)+\alpha\right] \end{aligned} \tag{33-1}$$

where $E_0^{(1)}$ and $E_0^{(2)}$ are the amplitudes and α is a phase angle; see Figure 33-5. If α is a random variable, then the radiation is said to be *unpolarized*. If $\alpha = 0$ or π, or one of $E_0^{(1)}$ or $E_0^{(2)}$ vanishes, then it is said to be *plane* or *linearly polarized* along the direction of the total field. If $\alpha = \pm\pi/2$ and $E_0^{(1)} = E_0^{(2)}$, then we say that the wave is *circularly* polarized. For other values of α the light is characterized as being *elliptically* polarized.

Suppose that a beam of unpolarized light that is originally traveling in air strikes the interface of a medium of refractive index n. If the angle of incidence is *Brewster's angle* θ_B defined by

$$\tan \theta_B = n \tag{33-5}$$

then all of the reflected light is plane-polarized perpendicular to the plane of incidence.

The intensity I of an originally unpolarized light beam after being transmitted through a polarizer and an analyzer with axes at an angle θ relative to each other is

$$I = I_0 \cos^2 \theta \tag{33-7}$$

where I_0 is the maximum intensity observed for parallel axes. This relation is known as *Malus' law.*

If a linearly polarized wave is normally incident on a doubly refracting plate with optic axis as shown in Figure 33-17, the phase shift δ between the emerging components polarized along and perpendicular to the optic axis is

$$\delta = \frac{2\pi}{\lambda} d(n_o - n_e) \tag{33-14}$$

Here, λ is the free-space wavelength, n_o and n_e are the ordinary and extraordinary refractive indices, and d is the thickness of the plate.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: (a) polarized light; (b) birefringence; (c) Brewster's angle; (d) uniaxial crystal; and (e) dichroic crystal.
2. Define or describe briefly what is meant by the following terms: (a) optic axis; (b) axis of easy transmission; (c) principal axes; (d) quarter-wave plate; and (e) extraordinary ray.
3. Contrast and compare the following pairs of light rays: (a) linearly polarized–unpolarized; (b) circularly polarized–linearly polarized; and (c) circularly polarized–elliptically polarized.
4. If you look at the sky on a clear, sunny day through a Polaroid sheet, you will observe a variation in intensity as the Polaroid is rotated about an axis along the line of sight. What conclusions can you draw about the state of polarization of this light? How could you resolve any ambiguities?
5. Describe in physical terms the mechanism underlying the operation of a dichroic material such as Polaroid.
6. What physical characteristics of a doubly refracting crystal, such as quartz, bring about its unusual optical behavior? Does this mean that ordinary commercial cellophane, which is also birefringent, is crystalline? Explain.
7. Unpolarized light is normally incident on a thin sheet of material. Describe the state of polarization of the emerging beam if the sheet is: (a) Polaroid; (b) birefringent with the optic axis in the plane of the sheet; (c) a quarter-wave plate.

8. Would you expect that the axis of easy transmission for the lenses of Polaroid sunglasses to be vertical or horizontal when worn by a person? (*Hint*: How could you eliminate some of the rays reflected from water (refer to Figure 33-8)?)
9. Two parallel Polaroid strips have their axes at right angles so that no light comes through the combination. Suppose that a third Polaroid strip is placed between these two. Explain in detail why light may now get through this combination of three Polaroid strips.
10. *Unpolarized* light is transmitted through a quarter-wave plate. What is the state of polarization of the emerging beam? Would your answer also hold for a half-wave plate?
11. What is the final polarization state of an originally right circularly polarized light beam after going through a half-wave plate? Justify your answer.
12. Linearly polarized light is viewed through a quarter-wave plate. What is the nature of the observed beam? Does the intensity vary as the quarter-wave plate is rotated about the direction of the incident beam? Explain.
13. What are the largest and smallest values for the speed of the e-ray in **(a)** quartz? **(b)** calcite? Use the data in Table 33-1 and Figure 33-15.
14. Which travels faster in a positive crystal, the o-ray or the e-ray? In a negative crystal?
15. An o-ray travels along the x-axis of a coordinate system fixed relative to a uniaxial crystal whose optic axis is along the y-axis. What is its direction of polarization? What would be the direction of polarization if it were an e-ray?
16. What is the final state of polarization of an initially plane-polarized light beam after being transmitted successively through a quarter-wave plate and then a half-wave plate? Does the angle between the two optic axes play a role in your answer? Explain.
17. In Figure 33-13 determine the general direction of the optic axis in the crystal. (*Hint*: Refer to Figure 33-16d.)
18. Elliptically polarized light is viewed through Polaroid. What is the state of polarization of the transmitted light? Is any variation in intensity observed as the Polaroid is rotated about the beam direction?
19. For the situation in Example 33-4 suppose that a quarter-wave plate is placed in front of the Polaroid. Describe the nature of the variation in intensity that is now observed as the Polaroid sheet alone is rotated about the beam axis. If the quarter-wave plate and the Polaroid sheet are rotated in unison, what is the observed variation in intensity?
20. Repeat Question 19, but suppose this time that the incident light is elliptically polarized.
21. Describe in what ways two quarter-wave plates, one made of calcite and one made of quartz, differ. What is the distinction between the states of polarization of linearly polarized light transmitted through each plate?
22. Describe the significance and usefulness of the quarter-wave and half-wave plates for producing and analyzing polarized light.

PROBLEMS

Unless a statement is made explicitly to the contrary, assume in the following that the component of a light beam with the E-vector along the axis of easy transmission undergoes no losses on traversing a sheet of Polaroid.

1. Calculate Brewster's angle for heavy flint glass, which has a refractive index of 1.65 for the sodium line at $\lambda = 5893$ Å.
2. At an angle of incidence of 63°, it is found that the light reflected from a certain transparent substance is plane-polarized perpendicular to the plane of incidence. What is the refractive index for the given substance?
3. An unpolarized light beam traveling in a medium of refractive index n_1 strikes the interface with a second medium of index n_2. Show that Brewster's angle θ_B for this case is

$$\tan \theta_B = \frac{n_2}{n_1}$$

4. If θ_c is the critical angle for a certain refracting medium when in air, and if θ_B is the value for Brewster's angle for the same medium in air, show that

$$\cot \theta_B = \sin \theta_c$$

5. A monochromatic unpolarized light beam of intensity of 0.2 W/m^2 is normally incident onto a Polaroid sheet. **(a)** What is the intensity of the transmitted light? **(b)** What is the intensity observed if this beam goes through a second Polaroid sheet with its axis of easy transmission at an angle of 30° with respect to the first?
6. Let I_0 be the intensity observed for a light beam transmitted through two Polaroid sheets with parallel axes of easy transmission. If a third sheet is put between the original two in such a way that the observed intensity is reduced to $0.3I_0$, at what angle is the axis of easy transmission of the third sheet relative to that of the others?
7. A monochromatic wave of wavelength λ travels along the positive z-axis of a certain coordinate system. Identify the state of polarization in each case if the electric field in the wave is given by:

(a) $\mathbf{E} = \mathbf{i}E_0 \sin \frac{2\pi}{\lambda}(z - ct)$

(b) $\mathbf{E} = \mathbf{i}E_0 \sin \frac{2\pi}{\lambda}(z - ct) + 2\mathbf{j}E_0 \sin \frac{2\pi}{\lambda}(z - ct)$

(c) $\mathbf{E} = \mathbf{i}E_0 \sin \frac{2\pi}{\lambda}(z - ct) + \mathbf{j}E_0 \cos \frac{2\pi}{\lambda}(z - ct)$

(d) $\mathbf{E} = \mathbf{i}E_0 \sin \frac{2\pi}{\lambda}(z - ct) - 2\mathbf{j}E_0 \cos \frac{2\pi}{\lambda}(z - ct)$

(e) $\mathbf{E} = \mathbf{i}E_0 \sin \frac{2\pi}{\lambda}(z - ct) + \mathbf{j}E_0 \sin \left[\frac{2\pi}{\lambda}(z - ct) + \frac{\pi}{3}\right]$

8. Consider a rotation of the coordinate system in Problem 7 by an angle α about the z-axis. The unit vectors $\mathbf{i}'$ and $\mathbf{j}'$ along the *new* axes are related to the old ones by

$$\mathbf{i}' = \mathbf{i} \cos \alpha + \mathbf{j} \sin \alpha$$
$$\mathbf{j}' = -\mathbf{i} \sin \alpha + \mathbf{j} \cos \alpha$$

(a) Express the electric fields in (a) and (b) of Problem 7 in terms of these unit vectors.

(b) Prove that the field in (c) of Problem 7 represents circularly polarized light in terms of the new axes for any choice for the angle α.

9. An unpolarized beam of light is incident on three parallel sheets of Polaroid, each of whose axes is at an angle of 45° with respect to that of its predecessor.

(a) What fraction of the incident radiation intensity comes through?
(b) If the middle Polaroid sheet is removed, what fraction of the incident intensity comes through?
(c) Compare your answers to (a) and (b), and explain them in physical terms.

10. Three strips of Polaroid are parallel and lined up so that the maximum intensity I_0 is transmitted.
(a) If the middle one is rotated by 30°, what is the intensity now?
(b) If in the initial situation, the last one is rotated by 30°, what intensity gets through?
(c) What is the intensity if the middle and the last Polaroid strip are rotated simultaneously by 45° in the same direction?

11. By starting with (33-3) and an analogous formula for left circularly polarized light (with the same amplitude and phase) prove that a plane-polarized light beam may be thought of as the superposition of two circularly polarized light beams of the same amplitude and phase, one polarized to the left and the other to the right.

12. Imagine viewing, through a Polaroid sheet, an incident light beam whose electric vector is characterized by the electric field in Problem 7b. What is the ratio of the minimum to the maximum transmitted intensity observed as the Polaroid is rotated about the beam axis?

13. Repeat Problem 12, but this time suppose the electric field in the incident beam is that given in Problem 7c. Explain why your answer *must* differ from that in Problem 12.

*14. Suppose that in Figure 33-16c the unpolarized light beam were incident at an angle θ. Assuming the optic axis of the crystal to be perpendicular to the plane of incidence:
(a) Why is Snell's law applicable to both the *o*-ray and the *e*-ray in this case?
(b) Assuming further that the crystal is calcite, and that $\theta = 30°$, use the data in Table 33-1 to calculate the angular separation between the two rays in the crystal.

15. Making use of (33-13), calculate by use of the data in Table 33-1 the velocity of an *e*-ray traveling at an angle of 30° with respect to the optic axis in (a) calcite and (b) quartz.

16. Show by use of (33-12) that the speed u of an *e*-ray which travels in a uniaxial crystal along a direction making an angle θ with the optic axis is

$$u = \frac{c}{\sqrt{n_o^2 \cos^2\theta + n_e^2 \sin^2\theta}}$$

17. Making use of the result of Problem 16, find the maximum and minimum values for the velocity of an *e*-ray in (a) calcite and (b) quartz.

18. Figure 33-19 shows an unpolarized light beam normally incident on what is known as a "Rochon" prism made of quartz. As shown, the light on entering travels along the optic axis of the first prism and then is doubly refracted at the interface with the second prism, whose optic axis is perpendicular to the plane of incidence.

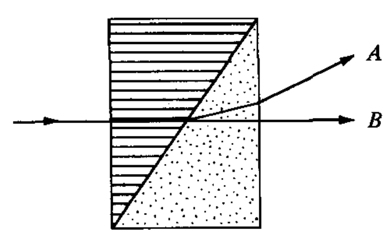


Figure 33-19

(a) Why is ray B undeviated?
(b) Which of the two rays is the e-ray?
(c) Describe the state of polarization of the two emerging rays.

*19. In Figure 33-19, if the incident ray strikes the interface between the two prisms at an angle of incidence of 30°, calculate the angular separation between the two emerging rays. Assume that both prisms are made of quartz and use the data in Table 33-1.

*20. Figure 33-20 shows an unpolarized light beam normally incident on a "Wollaston" prism. As shown, the incident beam on entering the prism first travels perpendicularly to the optic axis and then is doubly refracted at the interface with a second prism whose optic axis is perpendicular to the plane of incidence. Assuming that both prisms are made of quartz, and using the data in Table 33-1:

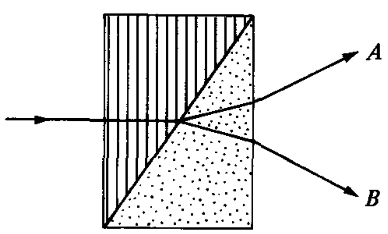


Figure 33-20

(a) Which of A and B is the o-ray and which is the e-ray?
(b) What is the direction of polarization of each of these rays?
(c) Assuming an angle of incidence of 30° at the interface, calculate the angular separation of the o-ray from the incident direction.

21. Repeat Problem 20, but suppose that this time the prisms are made of calcite.

22. (a) Show that the minimum thickness d_0 of a quarter-wave plate is given by

$$d_0 = \frac{\lambda}{4|n_e - n_o|}$$

where λ is the free-space wavelength of the light being used and n_e and n_o the two refractive indices.
(b) What other values can the thickness d have so that it is still a quarter-wave plate?
(c) Calculate the smallest thickness of a quarter-wave plate made of calcite for yellow light of wavelength 5893 Å.

23. Repeat all parts of Problem 22 for a half-wave plate.

24. By use of the data in Table 33-1 calculate the ratio of the thicknesses of two quarter-wave plates, one made of calcite and the other made of quartz. Assume the minimum thickness in each case.

25. A light beam, plane-polarized along the vertical, is transmitted through a quarter-wave plate with its optic axis at an angle θ with respect to the vertical. Assuming that $0 < \theta < 45°$:
(a) Show that the light is elliptically polarized.
(b) What is the spatial orientation of the major axis of the ellipse?
(c) Show that the ratio of the semimajor axis to the semiminor axis of the ellipse is cot θ.

26. Prove that if linearly polarized light is normally incident on a half-wave plate, and if the polarization direction makes an angle θ with the optic axis of the plate, then the plane of polarization of the transmitted light makes an angle 2θ with respect to direction of the incident polarization.

*27. Prove the following:
(a) If transmitted through Polaroid, circularly polarized light comes out plane-polarized.
(b) If transmitted through Polaroid,

elliptically polarized light comes out plane-polarized.

(c) If the Polaroid in (a) is rotated about the direction of the beam, no variation in intensity is detected.

(d) If the Polaroid in (b) is rotated about the direction of the beam, there is a variation in intensity but some light is always transmitted.

28. A quarter-wave plate is designed to be used at a wavelength of 4000 Å. Suppose that in an experiment plane-polarized light of wavelength 5000 Å is transmitted through it. If θ is the angle between the optic axis of the plate and the polarization direction of the incident beam, describe the polarization state of the emerging light if: **(a)** $\theta = 0$; **(b)** $\theta = 90°$; **(c)** $\theta = 45°$. Is there any value for θ for which the light is circularly polarized?

29. For the physical situation in Figure 33-18, suppose that the axes of both Polaroid sheets are vertical. What is the variation of the transmitted intensity observed in terms of θ as the half-wave plate is rotated about the direction of the beam?

30. If the axis of the polaroid strip next to the observer in Figure 33-18 is rotated so that its optic axis makes an angle α with respect to the vertical (the other one still being along the vertical), find the variation in the observed intensity as a function of the angles θ and α as the half-wave plate is rotated about the direction of the incident beam.

***31.** You have available only a Polaroid strip and a quarter-wave plate with which to analyze the state of polarization of a given beam of light. Assuming that the light beam is either plane-polarized, circularly polarized, or elliptically polarized, set up a procedure to determine which it is.

***32.** Repeat Problem 31, but this time suppose that the light beam can be a mixture of any two of these three states of polarization.

***33.** Repeat Problem 31, but this time suppose that the incident beam could in addition be unpolarized.

34 Epilogue

An intelligence knowing, at a given instant of time, all forces acting in nature, as well as the momentary positions and velocities of all things of which the universe consists, would be able to comprehend the motions of the largest bodies of the world and those of the lightest atoms... provided his intellect were sufficiently powerful to subject all data to analysis; to him nothing would be uncertain; both present and future would be in his eyes.

PIERRE S. DE LAPLACE (1749–1827)

34-1 Introduction

It is not very difficult for us today to see why the late nineteenth-century physicists might have been overly proud of the achievements in their discipline during the preceding two centuries. With Newton's mechanics and Maxwell's theory of electromagnetism firmly established, it appeared to them to be only a matter of time until all physical happenings in the world could be described in full detail. They knew, for example, that mechanics, as encompassed in Newton's laws of motion and of gravity, correctly described not only the large-scale motions of the solar system and of the galaxy, but those of bodies of ordinary size as well. Moreover, as they had seen demonstrated by developments in the kinetic theory of gases and statistical mechanics, these laws appeared to be applicable on a microscopic level as

well. Similarly, thanks to Maxwell they had acquired insights into electromagnetic phenomena—which by this time included optics as a special case—that no one could have conceived of only 50 years earlier.

As viewed by these late nineteenth-century physicists, then, these basic physical laws—or laws of *classical physics*, as they are known today—were logically consistent, complete, and fully capable of accounting in detail for all phenomena in the physical universe. Indeed, many expressed sorrow and concern for the plight of the twentieth-century physicist, whose central problem, they feared, would amount principally to ascertaining the "next decimal place." As Maxwell himself expressed it at the inauguration of the Devonshire (Cavendish) Laboratory at Cambridge in 1871:

> . . . The opinion seemed to have got abroad that in a few years all the great physical constants will have been approximately estimated and that the only occupation which will then be left to men of science will be to carry on these measurements to another place of decimals.

As subsequent events have shown, this sort of complacency has no place in scientific research and thought. Indeed, the carrying out of measurements to the "next decimal place" plays the extremely important role in scientific research of helping to ascertain the limits of validity of physical laws. Except for the much rarer discovery of an altogether unforeseen phenomenon, this is often the only means available for uncovering new physical laws. In any event, largely in the pursuit of this secondary objective, during the 30-year period beginning about 1895 many physicists were pursuing studies to see to what extent the laws of classical physics were applicable at the microscopic level. Much to their surprise, they found that these laws, which had worked so well for macroscopic systems, were simply not applicable, without modification, to individual atoms and molecules. That is, they found that the application of the laws of Newtonian mechanics, and Maxwell's theory to atoms and molecules led in some cases to both quantitative as well as qualitative contradictions with experiment.

Out of this period of intellectual ferment there emerged ultimately two entirely new physical theories. These are (1) the *special*[1] *theory of relativity*, which was proposed by Albert Einstein in 1905; and (2) the theory of *quantum mechanics*, which was proposed by L. de Broglie, E. Schrödinger, W. Heisenberg, and others during the period 1924-1926.

With the theory of relativity, as we have seen earlier, Einstein extended the domain of validity of classical physics to high-speed phenomena. He argued that the laws of physics must be invariant, not under Galilean transformations as had been assumed in Newtonian mechanics, but rather under Lorentz transformations. As we saw in Sections 3-12 and 9-10, these

[1]The general theory of relativity, which was also proposed by Einstein in 1916, plays an important role only for phenomena that take place near very massive bodies, such as very large stars. It is normally of no significance in considerations of terrestrial, microscopic phenomena, and hence will not be considered further here.

ideas led to certain modifications of a number of physical quantities, including, for example, the mass, the energy, and the momentum of a particle. It is significant, however, that all the basic features of Newton's laws of motion, such as their deterministic nature and their associated laws of energy and momentum conservation, were retained.

The second theory developed during this period is quantum mechanics. In simplest terms this theory may be viewed as the generalization of Newton's laws required to describe systems with masses and dimensions of the order of those of atoms and molecules. The predictions of quantum mechanics for macroscopic bodies are precisely the same as those of Newtonian mechanics. Hence, quantum mechanics is often characterized as the generalization of Newtonian mechanics required to describe the behavior of microscopic systems. However, the changes imposed by the theory of quantum mechanics on basic classical concepts are much more severe, from a philosophical point of view, than those imposed by the theory of relativity. In particular, the deterministic nature of Newtonian mechanics, as expressed in the quotation from Laplace at the beginning of this chapter, had to be forfeited. Quantum mechanics is not a deterministic theory in the sense we are used to in classical mechanics.

34-2 Some experimental difficulties

By way of an introduction to some of the difficulties encountered in applying the laws of classical physics to microscopic phenomena, in this section we shall list certain predictions made by the classical laws. With each of these predictions we shall also state an experimental result that is incompatible with it. The remaining sections of this chapter are then devoted to a discussion of some of the assumptions—which eventually led to the discovery of quantum mechanics—that were made in order to account for these experimental results. This list is representative, but in no sense all-inclusive.

1. **Classical prediction:** *An accelerating charged particle radiates electromagnetic energy.*

This prediction is in conflict with the self-evident fact that atoms are stable. An atom, as we know, consists of a positively charged nucleus, about which orbit an appropriate number of negatively charged electrons, so that overall the atom is electrically neutral. In order for this picture to be viable, it is necessary, according to classical ideas, for these electrons to accelerate as they orbit about the nucleus. But this means (Section 29-9) that they must radiate electromagnetic energy and thus gradually spiral into the nucleus. It follows therefore that the very existence of atoms is in contradiction with one or more facets of classical physics.

2. **Classical prediction:** *The energy of a mechanical system can assume any value from a continuous range of values.*

This contradicts the fact that the energy of bound microscopic systems, such as atoms, molecules, and nuclei generally can assume only certain discrete values; other values are forbidden. The allowable energy values for these systems are known as *energy levels* or *energy eigenvalues.*

3. **Classical prediction:** *An electromagnetic wave is a wave and not a stream of corpuscles such as envisioned by Newton.*

This conflicts with the fact that although, under many conditions, electromagnetic radiation is indeed a wave phenomenon, there exist others for which it is best described as consisting of particulate bundles of energy now known as *photons.* In other words, electromagnetic radiation has a dual nature: under some circumstances it acts as a wave and under others as a stream of particles. The relation between the parameters characterizing these two aspects is given in (30-1), which expresses the energy E and the momentum p of the photons in the wave in terms of its wavelength or frequency:

$$E = h\nu \qquad p = \frac{h}{\lambda} \tag{30-1}$$

4. **Classical prediction:** *Particles, such as electrons or atoms, do not exhibit wavelike characteristics.* In particular, if a stream of electrons is incident on a slit system, such as that in Figure 34-1, neither diffraction nor interference phenomena will be observed on the screen.

This is in conflict with the experimental fact that if a beam of electrons is incident from the left on the two-slit system in Figure 34-1, then the electrons striking the screen will, under appropriate conditions, exhibit diffraction and interference. Specifically, the observed intensity with both slits open will not, in general, be the same as the sum of the intensities observed if only one slit is open at a time. This is a characteristic feature of

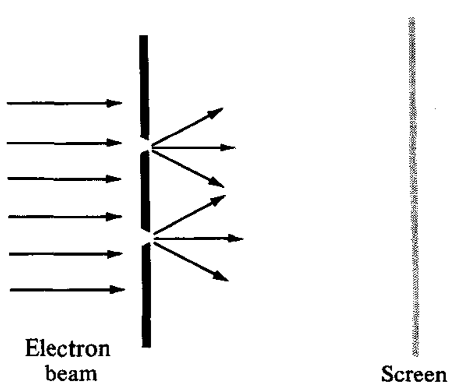


Figure 34-1

an interference effect and demonstrates that matter—the electron beam in this case—has wavelike properties. Hence, just as there are particulate aspects associated with electromagnetic waves, experiments of this type show that there are also wavelike aspects associated with matter. The very existence of these *matter waves*, as they are known, makes it unambiguous that Newtonian mechanics is not always applicable at the microscopic level.

34-3 Thermal radiation

Experiment shows that if a noninflammable body, such as an iron poker, is heated, it will emit radiation whose quantity and quality generally depend on the temperature. At low temperatures, the wavelengths of the emitted rays are mainly in the infrared region of the electromagnetic spectrum (Section 29-7) and thus not directly observable by the unaided eye. However, as the temperature is raised, the body begins to glow with a dull red color which gradually turns a bright red. Eventually, when its temperature is sufficiently high, the radiated light appears to be white and the body is then said to be "white hot." The glow of the tungsten filament in a light bulb is an example of such a white-hot radiator.

A spectroscopic investigation of this type of radiation shows that it consists of a mixture of light of various wavelengths distributed continuously throughout the entire visible spectrum. That is, all colors, from red at one end to deep violet at the other, are present in the radiation. Moreover, the use of infrared and ultraviolet sensors shows that the emitted radiation is not confined to visible wavelengths alone but comes with varying intensities from neighboring parts of the spectrum as well. Thus, at low temperatures when the radiating body appears dull red, the most intense visible radiation it emits is at wavelengths of the order of 6500 Å; but there is a significant amount radiated at the invisible infrared wavelengths as well. As the temperature is raised, the maximum in intensity moves continuously to shorter and shorter wavelengths, as a result of which ultraviolet light becomes a more and more significant component of this radiation.

As a general rule, the amount and quality of the radiation emitted by a body depend on the nature of the body itself. However, under conditions of thermal equilibrium—that is, when the energy radiated per unit area per unit time by the body is equal to that absorbed from external sources—the nature and quality of this radiation must be the same for all bodies. For, how else could a state of thermal equilibrium exist? It follows that under conditions of thermal equilibrium, *all* bodies at temperature T will radiate electromagnetic energy with the same intensity at corresponding wavelengths. This energy radiated under conditions of thermal equilibrium is known as *blackbody radiation* or *isothermal radiation*. It depends only on the temperature and not on the nature of the material doing the radiating.

A convenient way to measure the distribution in wavelength of blackbody

radiation is by use of a hollow body or cavity whose walls are maintained at some fixed temperature T. The radiation emitted by the interior walls of the cavity will be trapped inside, and eventually the enclosed radiation will come into thermal equilibrium with the walls. The characteristic features of blackbody radiation can then be measured by making a small opening in the enclosure and observing the emerging radiation. Figure 34-2 is a plot of the logarithm of the blackbody radiation intensity I or the energy radiated per unit time per unit area in a wavelength interval $d\lambda$ as a function of wavelength λ for various temperatures. The total energy radiated per unit area per unit time at all wavelengths is proportional to the area under the appropriate curve. Note that, consistent with our expectations, the higher the temperature the more energy is radiated.

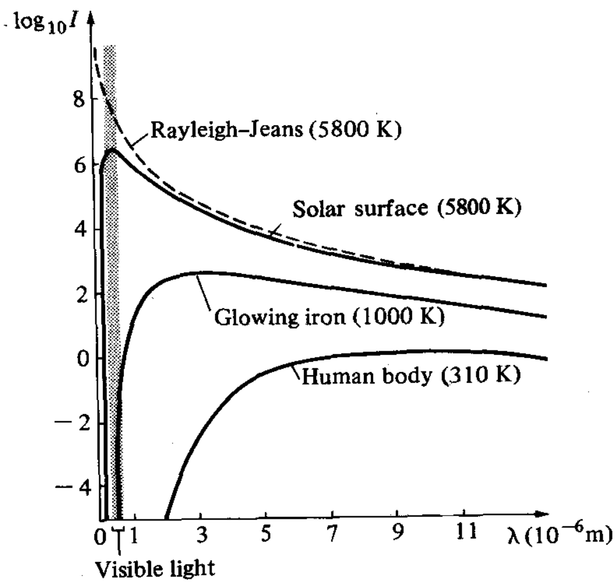


Figure 34-2

Isothermal radiation of this type plays an extremely important role in the physical world. It characterizes, for example, the heat radiation, generally at infrared wavelengths, radiated and absorbed by our own bodies at 310 K. Similarly the radiation emitted from the surface of the sun, which is at a temperature of approximately 5800 K, is also of this type. Indeed, the surface temperature of many stars is often inferred by studying the nature of the radiation they emit. Evidently, the continuous isothermal radiation emitted by a body is a very important physical quantity and one with which any acceptable physical theory must be able to come to grips.

34-4 Planck's proposal

If we were to apply the ideas of classical physics to the problem of blackbody radiation we would find a surprisingly simple and unambiguous

answer. This is the *Rayleigh–Jeans* formula:

$$I = \frac{2\pi ckT}{\lambda^4} \tag{34-1}$$

where $I \equiv I(\lambda, T)$ is the energy radiated per unit area, per unit time, and per unit wavelength interval at the wavelength λ when the body is in thermal equilibrium at the absolute temperature T. The constants k and c here are respectively, Boltzmann's constant and the speed of light, and have the values $k = 1.38 \times 10^{-23}$ J/K and $c = 3.0 \times 10^8$ m/s.

Unfortunately, this very simple Rayleigh-Jeans formula in (34-1) has a very serious drawback. For although it is in excellent, actually perfect, agreement with experiment at long wavelengths, it is in violent disagreement at short wavelengths. Indeed, because of the λ^{-4} singularity at $\lambda = 0$, the integrated spectrum—that is, the total energy radiated at all wavelengths—is infinite! This implies that at least one of the laws of classical physics that was used in deriving (34-1) and was assumed to be applicable at all wavelengths must be seriously in error. There is no ambiguity. As evidenced by the classical (dashed) curve in Figure 34-2, the classical formula disagrees with experimental facts.

In 1900 Max Planck (1858–1947) proposed a new formula for this important physical quantity I. Planck's formula is

$$I = \frac{2\pi hc^2}{\lambda^5}\left\{\frac{1}{\exp\left[hc/\lambda kT\right] - 1}\right\} \tag{34-2}$$

where k, c, λ, and T are as defined above and $h \cong 6.63 \times 10^{-34}$ J-s is Planck's constant. The various curves in Figure 34-2 are actually plots of this formula. For long wavelengths, it is easy to show by expanding the exponential that (34-2) reduces to the Rayleigh–Jeans formula in (34-1). By contrast to the incorrect Rayleigh–Jeans result at short wavelengths, however, Planck's formula for I is consistent with experiment at all wavelengths. Its prediction of an exponential decrease with decreasing wavelengths is fully in accord with observations at all wavelengths.

Now how did Planck arrive at this formula? Evidently an assumption that goes beyond the laws of classical physics is required. Indeed, to obtain (34-2), Planck was forced to make the hypothesis that only discrete units of radiant energy could be emitted by a body in thermal equilibrium. In other words, he found it necessary to assume that there is a smallest unit or *quantum* of electromagnetic energy that the constituent "oscillators" or atoms of a radiating body can emit. The result in (34-2) then followed provided that this quantum of energy, ϵ, is related to the frequency ν of the oscillators by

$$\epsilon = h\nu \tag{34-3}$$

where h is Planck's constant.

This hypothesis in (34-3) that the energy emitted by the constituents of a body in thermal equilibrium can occur only in discrete bundles or quanta is

completely foreign to classical physics. It states, in effect, that there is a smallest unit of energy, $h\nu$, and that all energy exchange can occur only in integral multiples of this unit. However, as strange as this hypothesis may sound at first, no one can quarrel with the success of its prediction, at least that in (34-2).

34-5 Quantization of the electromagnetic field

In a classic paper written in 1905, Einstein first suggested extending Planck's ideas on quantization to include the electromagnetic field itself. During the same year that he enunciated the theory of relativity he made this proposal in connection with an attempt to explain a phenomenon then under active experimental study and known as the *photoelectric effect*. In essence, he put forward the idea that an electromagnetic wave of frequency ν could, under some circumstances, be considered as a stream of corpuscles, or *photons* as we call them today, each with energy ϵ given in (34-3). Let us consider this matter briefly.

Experiment shows that if an electromagnetic wave of an appropriate wavelength is incident on certain substances, for example, zinc, potassium, or copper, electrons are emitted from its surface. These ejected electrons are called *photoelectrons*, and the phenomenon itself is known as the photoelectric effect. Figure 34-3 shows schematically an experimental setup for measuring the characteristics of these emitted electrons. An evacuated tube A contains a sample B and a collector C, which are maintained at some potential difference by a battery of emf $\mathscr{E}$. Under ordinary circumstances, with no radiation incident on B, the galvanometer reading is zero, since no current can flow around an open circuit. However, if light of an appropriate frequency is shined on the sample then, as shown in the figure, a current is found to flow through the galvanometer. The direction of this current corresponds to a stream of electrons going from B to the collector C inside the tube.

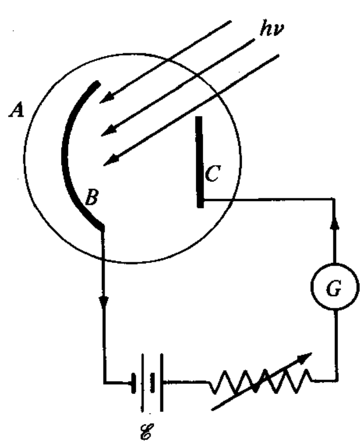


Figure 34-3

Studies of this effect usually involve measuring the current through the galvanometer as a function of the frequency and intensity of the incident radiation. These studies show that the photoelectric effect has a number of features that are unusual from the classical physicist's point of view. Very striking among these are:

1. Regardless of the intensity of the incident radiation, no electrons are emitted if the frequency is below a threshold value ν_c, which in general is different for different materials.
2. If the frequency of the incident radiation exceeds the threshold value ν_c for the given material, then electrons are emitted. The number of these electrons is proportional to the intensity of the incident light.
3. The distribution of the kinetic energies of the ejected electrons does not depend on the intensity of the incident radiation.
4. The maximum kinetic energy K_{max} of the emitted electrons is greater, the greater is the frequency ν of the incident radiation.

Each one of these properties of the photoelectric current is difficult to understand in terms of the concepts of classical physics. For example, the classical physicist would expect that the kinetic energies of the electrons would depend on the intensity of the incident radiation and certainly not its frequency. (See, for example, (29-25) in this connection.) The fact that below a threshold frequency ν_c there is no photoelectric current regardless of the incident intensity would be equally incomprehensible to him.

In a classic paper published in 1905, Einstein showed that these anomalous features could be accounted for by extending Planck's ideas of quantization to the electromagnetic field itself. He postulated that an electromagnetic wave of frequency ν consists of a stream of corpuscles—or *photons* as they are called today—each with energy E given by

$$E = h\nu \tag{34-4}$$

with h the same Planck's constant as in (34-3). Assuming, further, that a given electron in the metal is not likely to absorb more than one of these photons, he argued that the maximum energy acquired by any electron must be $h\nu$. Therefore, if this energy $h\nu$ is not large enough for an electron to overcome the potential barrier V_0 at the surface of the metal, no electrons will be emitted. However, if the energy $h\nu$ of the incident photons exceeds eV_0, then it *is* energetically possible for the electrons to leave the surface. By conservation of energy, the maximum kinetic energy of these ejected electrons must be the difference $(h\nu - eV_0)$. Hence

$$K_{max} = h\nu - eV_0 \tag{34-5}$$

This is Einstein's famous photoelectric-effect equation. The quantity eV_0 here is also known as the *photoelectric work function*. For a typical metal, such as zinc, it has a value of the order of 3.0 eV.

The value of K_{max} associated with a fixed frequency can be conveniently

measured by reversing the polarity of the battery in Figure 34-3 and determining the minimum potential difference between B and C for which no current flows. Figure 34-4 shows a plot of K_{max} against ν as obtained in this way. The fact that the curve is a straight line is a beautiful confirmation of Einstein's proposal and its consequence in (34-5). The intercept along the horizontal axis is the threshold frequency ν_c and is generally different for different materials. It is related to the potential barrier V_0 by $\nu_c = eV_0/h$. Note that the slope of the curve is the same for all materials and has the value h.

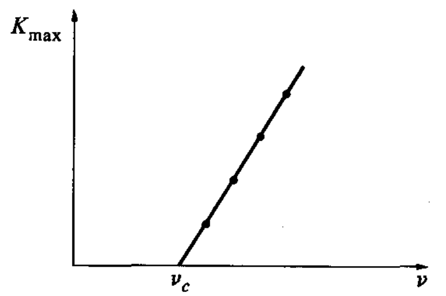


Figure 34-4

34-6 The Compton effect

A second, and equally impressive, experimental justification for the photon concept was discovered by Arthur H. Compton (1892–1962) about two decades later in 1923. Compton was interested in the distribution of the X rays scattered by the electrons in a carbon target. To this end he carried out experiments that measured the wavelength λ' of the emerging X rays as a function of the scattering angle θ and the wavelength λ of the incident X ray; see Figure 34-5. Assuming, with Einstein, that the incident X rays may be thought of as a stream of particles, each of energy $h\nu = hc/\lambda$, and with momentum $h\nu/c = h/\lambda$, and that energy and momentum are conserved during the collision, he deduced the relationship

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta) \tag{34-6}$$

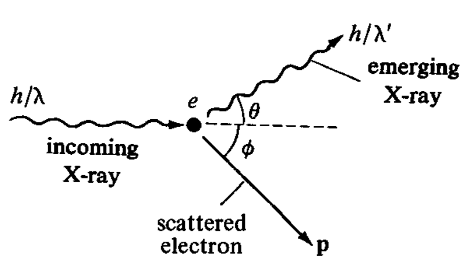


Figure 34-5

where m is the mass of an electron. And this is precisely the relation between λ' and θ that he measured in his experiment. The quantity h/mc in this relation has dimensions of length and is known as the *electron-Compton* wavelength. Its numerical value is approximately 2.4×10^{-12} meter. The term *Compton scattering* is used very generally today to describe processes that involve the scattering of photons from charged particles.

To derive (34-6), let us suppose that the velocity of the electron is sufficiently small that its momentum and kinetic energy are mv and $mv^2/2$, respectively. (See Problem 12 for the corresponding analysis for relativistic electrons.) Assuming, then, that before the collision the electron is at rest, and afterward it has a velocity v, it follows from the law of energy conservation that

$$\frac{hc}{\lambda} = \frac{hc}{\lambda'} + \frac{1}{2} mv^2 \tag{34-7}$$

since, for example, $h\nu = hc/\lambda$ is the initial energy of the photon. Moreover, since the momentum of a photon of energy $h\nu$ is $h\nu/c = h/\lambda$, it follows from the law of conservation of linear momentum that

$$\begin{aligned} \frac{h}{\lambda} &= \frac{h}{\lambda'} \cos\theta + mv \cos\phi \\ \frac{h}{\lambda'} \sin\theta &= mv \sin\phi \end{aligned} \tag{34-8}$$

where θ and ϕ are as defined in Figure 34-5. Finally eliminating the electron parameters v and ϕ from among (34-7) and (34-8), we find that the wavelength λ' of the scattered photon is indeed given by the very simple formula in (34-6).

34-7 Atomic spectra

Imagine an originally evacuated glass tube filled with an element in a gaseous state; for example, hydrogen, helium, or neon. If the gas is heated, say by passing an electric discharge through it, the tube will glow with a color characteristic of the given element. A spectroscopic analysis of this light shows that its spectrum consists of a sequence of discrete lines, each with its distinctive wavelength or color. Figure 34-6 shows schematically a possible experimental setup. Light emitted by the gas in the tube T is passed through a spectroscopic slit S and then viewed directly with the eye or on a viewing screen V after transmission through a diffraction grating D. Each line on the screen is an image of the spectroscopic slit associated with a particular wavelength. The sequence of lines obtained this way is called the *emission spectrum*, or the *line spectrum*, or simply the *spectrum* of the element in the tube. Experiment shows that no two elements have the same spectra; the emission spectrum of each element is a unique property

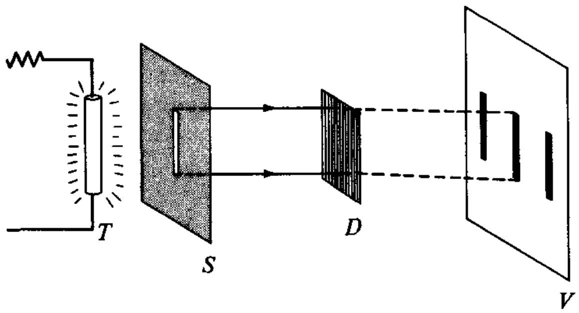


Figure 34-6

characteristic of the atoms of that element only. Figure 34-7 shows a photograph of the spectrum of copper.

Let us consider, for example, the spectrum of hydrogen. A spectroscopic analysis of this element shows the existence of four prominent lines in the visible region. These are a red-orange line at 6563 Å, a bluish-green line at 4861 Å, a blue-violet line at 4341 Å, and a deep violet line at 4101 Å. These four lines are members of a sequence of lines in the hydrogen spectrum known as the *Balmer series.* A second important series in the hydrogen spectrum, which occurs in the ultraviolet, is known as the *Lyman series.* Some members of this series are associated with the wavelengths 1216 Å, 1026 Å, and 973 Å. In a third series, known as the *Paschen series,* the lines occur in the infrared. Two examples are the lines at 18,756 Å and 12,821 Å.

Similarly, the emission spectrum of helium has seven prominent lines in the visible part of the spectrum. These include a red-orange line at 6678 Å, a yellow line at 5875 Å, and on to the violet end of the spectrum through the sequence of lines at 5015 Å, 4921 Å, 4713 Å, 4471 Å, and 4026 Å.

As a general rule, any element when in the gaseous state and heated sufficiently will emit radiation consisting of a collection of spectral lines characteristic of that element. Indeed, a practical way for ascertaining the existence of a certain element in a chemical substance is by heating it and then analyzing the spectrum of its vapor. Thus, the element sodium is known to emit two very closely spaced and intense yellow lines; one at 5890 Å and the other at 5896 Å. Hence, if these two lines are detected in the spectrum of some substance it is safe to conclude that the element sodium must be present in it. No other element emits these two lines.

Very closely related to the emission spectrum of an element is its *absorption spectrum.* This is obtained by making a spectroscopic analysis of continuous or thermal radiation after it has passed through a vapor of the

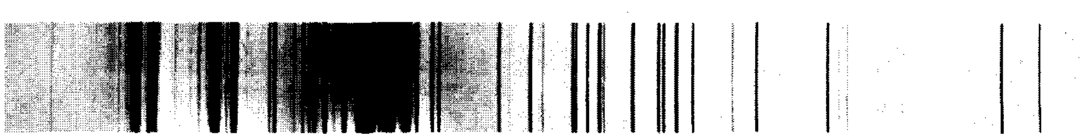


Figure 34-7 A photograph showing some of the spectral lines of copper. (Courtesy Bausch and Lomb, Analytical Systems Division.)

element in question. A possible experimental setup would be similar to that in Figure 34-6, but with the tube T replaced by a continuous source plus an intervening tube containing the vapor of the element. This time, as illustrated in Figure 34-8, the pattern on the screen is a continuous spectrum characteristic of thermal radiation, but cut by a series of dark lines. These lines constitute the absorption spectrum of the element. Assuming that the spectroscopic analysis is made by use of a diffracting grating, we may associate with each of these lines a wavelength in accordance with the basic formula in (32-18). We find in this way that each of the lines in the absorption spectrum coincides with some line in the emission spectrum. However, except under extreme conditions of gas temperature—for example, that of the gases surrounding the sun—*not* all lines of the emission spectrum have their counterparts in the absorption spectrum.

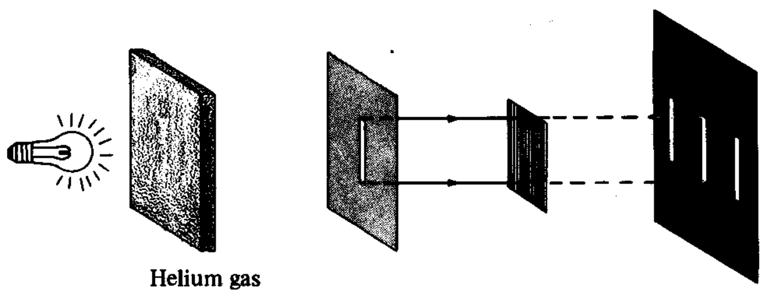


Figure 34-8

If, for example, a laboratory experiment is carried out to produce the absorption spectrum of hydrogen, in general only the Lyman series in the ultraviolet region is obtained. It is extremely difficult to produce the Balmer or the Paschen series in an absorption spectrum. To do so would require raising the gas temperature to prohibitively high values.

However, the fact that the Balmer and the Paschen lines do indeed exist in the absorption spectrum of hydrogen can be established by making a spectroscopic analysis of the radiation emitted by a star—for example, our own sun. Indeed, William H. Wollaston and Joseph Fraunhofer first discovered the existence of absorption spectra during the early part of the nineteenth century by making a spectral analysis of the light emitted by the sun. Fraunhofer is credited with the discovery of nearly 750 absorption lines in the solar spectrum. He was also the first to note that in many cases the dark lines in the absorption spectrum appear in the same positions as do certain emission lines of selected elements. Finally, in 1859 Gustav R. Kirchhoff established the significance of the Fraunhofer lines—as the lines in the solar absorption spectrum are known—by producing the absorption spectra of various elements in the laboratory. Today it is generally agreed that the surface of the sun is at a temperature of nearly 5800 K and radiates mainly isothermal radiation at that temperature. Prior to reaching the earth, however, this continuous spectrum of radiation passes through the hot gases

surrounding the sun. The dark lines observed in the solar spectrum arise because of the presence of a variety of elements in these vapors. Nearly 65 elements have been detected on the sun in this way. This method is also used to determine the chemical composition of other stars.

Although in this discussion we have considered only the spectra of atoms, it should be noted that other microscopic systems, including notably molecules and nuclei, also display discrete spectra. It is apparent, therefore, that these various spectra represent an important clue to the structure of all microscopic systems. And as such, the existence and quantitative features of these spectra should at least be consistent with the laws of physics. Yet, despite much effort, no one has ever succeeded in explaining even the existence of these spectra by use only of classical physics.

34-8 The Bohr model

The first successful quantitative explanation of atomic spectra was developed by Niels Bohr (1885–1963) in 1913. To achieve this, Bohr found it necessary to introduce a number of *ad hoc* assumptions, each one of which was in contradiction with, or went beyond some facet of, classical physics. Nevertheless, his model was extremely successful in accounting in detail for many of the significant spectroscopic features of the hydrogen atom, and thus it is of more than passing interest. Unfortunately, this *Bohr model*, as it is known, is applicable only to hydrogen, or hydrogenlike ions, such as He^{+}, Li^{++}, Be^{+++}, and so forth. It is not useful for describing the spectra of other atoms. Ultimately it was abandoned, but not until the discovery of quantum mechanics more than a decade later.

Because of the revolutionary nature of the postulates underlying Bohr's model and also because it leads to basically the same hydrogen spectrum as does the correct theory, let us consider this model briefly. For hydrogen Bohr's postulates may be stated as follows:

1. The electron travels about the proton in Keplerian orbits by virtue of the Coulomb force of attraction between them.
2. Not all orbits are allowed. The allowable ones are those for which the angular momentum L of the electron about the proton is an integral multiple of $h/2\pi$, where h is Planck's constant. The allowable orbits are therefore those for which the angular momentum L has one of the values:

$$L = \frac{nh}{2\pi} \qquad (n = 1, 2, \ldots) \tag{34-9}$$

   The quantity n is often referred to as a *quantum number.*
3. While in one of these allowed orbits the atom is stable. That is, the accelerating electron does *not* radiate electromagnetic energy, and its energy E_n in the orbit characterized by any integer n, is constant in time.

4. Radiation is emitted from the atom only when the electron makes a transition from one allowed orbit, say of energy E_n to another of lower energy $E_{n'}$. The frequency ν of the radiation emitted in such a transition is given by

$$h\nu = E_n - E_{n'} \tag{34-10}$$

Note the similarity between (34-10) and the Planck–Einstein energy quantization condition in (34-3) and (34-4).

In order to apply this theory to the calculation of the emission spectrum for hydrogen, let us consider an electron of mass m and charge $-e$ traveling at a uniform velocity $\mathbf{v}$ in a circular orbit of radius a about a proton of charge $+e$; see Figure 34-9. According to Coulomb's law, the electron experiences an attractive force $e^2/a^2 4\pi\epsilon_0$, and also undergoes a radially inward centripetal acceleration of magnitude v^2/a. Hence, by Newton's second law,

$$\frac{mv^2}{a} = \frac{e^2}{4\pi\epsilon_0}\frac{1}{a^2} \tag{34-11}$$

Since the angular momentum of the electron in this circular orbit is mva, it follows from Bohr's second postulate that for an allowable orbit the parameters v and a must be related by

$$mva = \frac{nh}{2\pi} \tag{34-12}$$

Finally, from the fact that the total energy E_n of the electron when in this orbit is the sum of its kinetic and potential energies, we have

$$E_n = \frac{1}{2}mv^2 - \frac{e^2}{4\pi\epsilon_0 a}$$

In order now to obtain an explicit formula for the electron energy in an allowable orbit, let us solve (34-11) and (34-12) for v and a and substitute the result into this formula for E_n. We find in this way Bohr's prediction for the energy levels for the hydrogen atom:

$$E_n = -\frac{1}{2}\frac{e^2}{4\pi\epsilon_0 a_0}\frac{1}{n^2} \qquad (n = 1, 2, \ldots) \tag{34-13}$$

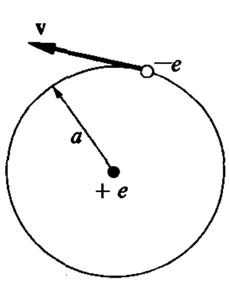


Figure 34-9

where a_0 is an important parameter known as the *Bohr radius*, and is defined by

$$a_0 = \frac{h^2\epsilon_0}{\pi m e^2} \tag{34-14}$$

Its numerical value is approximately 5.3×10^{-11} meter. According to Bohr's theory, then, the totality of values for the energy of an electron in a hydrogen atom are those given by (34-13). No other values for the energy are possible.

It is often very convenient and useful to represent the result in (34-13) on what is known as an *energy-level diagram*. As shown in Figure 34-10, such a diagram consists of an energy axis running vertically upward and a sequence of horizontal lines that intersect this axis at the allowable values for the energy. The state of lowest energy, or *ground state*, of the atom is represented by the lowest horizontal line on this diagram. As can be readily established by the substitution of the value $n = 1$ into (34-13), the energy E_1 of the ground state of hydrogen is -13.6 eV; the first *excited state* at $n = 2$ occurs at $-13.6\ \text{eV}/2^2 \cong -3.4$ eV, and so forth for the other excited states.

The vertical arrows in the energy-level diagram in Figure 34-10 refer to the possible transitions that the electron can make in going from a given excited state to another one of lower energy. The frequency ν associated with the light emitted in any one of these transitions can be easily calculated by use

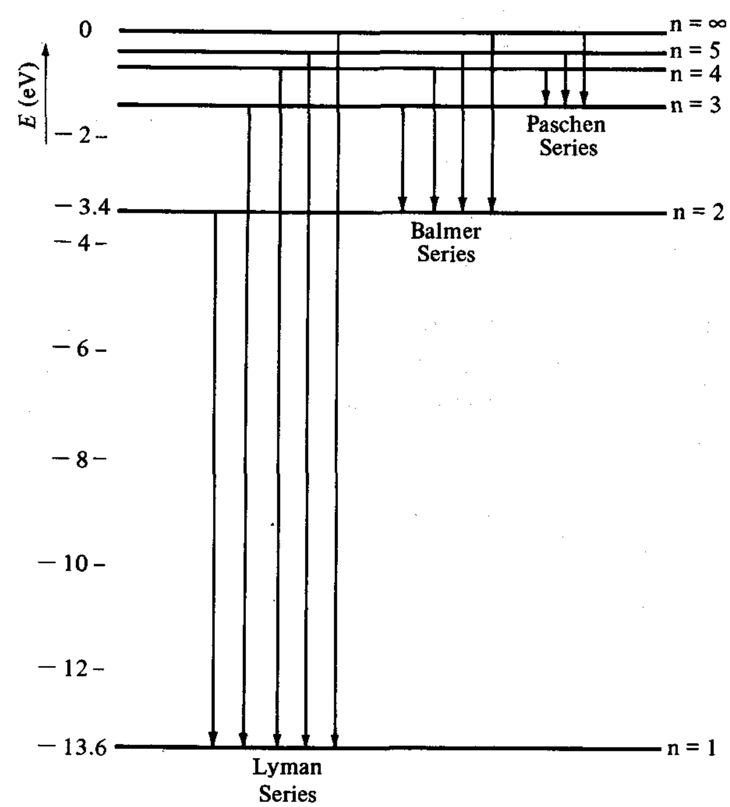


Figure 34-10

of (34-10) and (34-13). The result is given by the formula

$$\nu = \frac{1}{8h\pi\epsilon_0}\frac{e^2}{a_0}\left(\frac{1}{n'^2} - \frac{1}{n^2}\right)$$

where the quantum number n characterizes the initial state and n' that of the final state. As shown in the figure, for $n' = 1$ the frequencies for $n = 2, 3, \ldots$ correspond to the Lyman series. Similarly, for $n' = 2$ the transitions $n = 3, 4, \ldots$ represent the Balmer series, while for $n' = 3$ the lines for $n = 4, 5, \ldots$ are the Paschen series. In each case the spectral lines form an infinite sequence whose "series limit" is determined by setting $n \to \infty$. For example, since $\lambda = c/\nu$, it follows that the wavelengths of the lines of the Lyman series are given by

$$\frac{1}{\lambda} = \frac{1}{8\pi ch\epsilon_0}\frac{e^2}{a_0}\left(1 - \frac{1}{n^2}\right) \qquad n = 2, 3, 4, \ldots \qquad \text{(Lyman)}$$

and, consistent with experiment, this leads to wavelengths of 1216 Å, 1026 Å, 973 Å, . . ., 912 Å for $n = 2, 3, 4, \ldots, \infty$, respectively. Similarly, for $n' = 2$, we obtain the wavelengths of the lines in the Balmer series

$$\frac{1}{\lambda} = \frac{1}{8\pi ch\epsilon_0}\frac{e^2}{a_0}\left(\frac{1}{4} - \frac{1}{n^2}\right) \qquad n = 3, 4, \ldots \qquad \text{(Balmer)}$$

and this formula correctly predicts the observed wavelengths of 6563 Å, 4861 Å, 4341 Å, . . ., 3650 Å for $n = 3, 4, 5, \ldots, \infty$, respectively.

Thus by use of his model Bohr produced a quantitative way for describing the important features of the emission spectrum for hydrogen. Also, from the assumption that at low temperatures hydrogen atoms are mostly in the ground state and that on being exposed to radiation they preferentially absorb photons with energy $h\nu$ given by (34-10), it follows that the absorption spectrum should consist only of the Lyman series. This is also consistent with experiment. Thus, provided that we grant the validity of Bohr's postulates, the beginnings of a theory of atomic spectra emerges. Unfortunately, however, Bohr's theory leads to incorrect predictions for the spectra of atoms other than hydrogen and hydrogenlike atoms. Thus to obtain a realistic understanding of the spectra of atoms, as well as those of molecules and nuclei, it was necessary to await the discovery and development of quantum mechanics.

34-9 Matter waves

In 1924 still another hypothesis was proposed: this one by L. de Broglie (1892–   ). At the time of its being put forward this hypothesis appeared to be even more revolutionary than that of Planck and Einstein on the wave-particle duality of photons and those of Bohr on the existence of energy levels in atoms. Although it is not related to experiment as directly as

are Bohr's postulates, de Broglie's hypothesis played a significant role in the subsequent development of quantum mechanics.

In brief, what de Broglie suggested was that the Planck–Einstein wave-particle duality of the electromagnetic field was also shared by electrons. That is, he hypothesized that under some circumstances electrons behave as ordinary particles while under others they take on the characteristics of waves. Moreover, he proposed arguments which suggested that the wavelength λ associated with matter waves should be related to the momentum $p = mv$ of the electron by

$$\lambda = \frac{h}{p} \tag{34-15}$$

with h Planck's constant. In essence, de Broglie's argument for this relation consisted of the requirement that (34-9) be satisfied and that the circumference of the electron's orbit in the hydrogen atom be an integral number of wavelengths. (See Problem 23.)

An immediate and very important consequence of de Broglie's proposal was that it led other physicists, notably Erwin Schrödinger (1887–1961), to think in terms of matter waves and to attempt to quantify this concept by setting up an appropriate equation for such a wave. Ultimately, this effort led to the development of the theory of quantum mechanics, a key element of which is a wave equation for matter waves known as the *Schrödinger equation.* The role that this equation plays in quantum mechanics is very similar to that played by Newton's laws of motion in classical mechanics. Some knowledge of this wave equation of Schrödinger is essential for an understanding of much of the physics developed during the twentieth century.

A second important consequence of de Broglie's hypothesis was to motivate physicists to generate and to look for matter waves in the laboratory. In 1928 C. J. Davisson and L. H. Germer of the United States and G. P. Thompson of Scotland succeeded in this effort. In the Davisson–Germer experiment electrons with energies of the order of 100 eV were incident on a nickel crystal. An analysis of the scattered electrons showed that their spatial distribution could best be described as the diffraction of plane waves by the regularly spaced atomic planes in the crystal. Moreover, the wavelength λ associated with these waves turned out to be precisely that given by the de Broglie relation in (34-15). The experiment of G. P. Thompson was similar, but this time the electron diffraction patterns resulted from an electron beam going through very thin gold foils. Figure 34-11 illustrates the types of diffraction patterns found in this way. The existence of matter waves, as strange as this might appear to the classical physicist, is definitely an experimental reality of the physical world. The fact that particles other than electrons also exhibit wavelike behavior was subsequently confirmed in 1931 by Estermann, Frish, and Stern. They succeeded in demonstrating the diffraction of helium atoms by scattering a monoenergetic beam of these atoms from a lithium fluoride crystal.

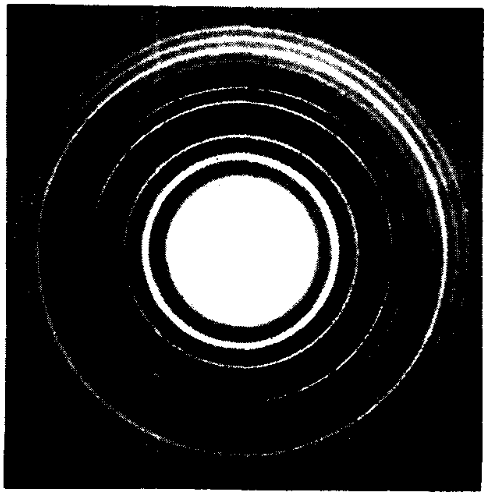


Figure 34-11 Diffraction pattern produced by directing an electron beam through polycrystalline aluminum. (Courtesy Project Physics, Holt, Rinehart and Winston, New York.)

It is important to recognize that the de Broglie relation in (34-15) is not in conflict with normal observations of macroscopic physical systems. As we saw in Chapter 32, the fact that light is a wave motion manifests itself only when it interacts with systems with dimensions of the order of its wavelength. For the case of a macroscopic body with momentum $p = mv \sim 1$ kg-m/s^2, the wavelength $\lambda = h/mv$ of the associated matter wave is, according to (34-15), of the order of 6×10^{-34} meter. Hence the wave aspects of this system are completely masked in this case. However, for the case of an electron with energy of 100 eV, as in the Davisson–Germer experiment, we find that $p \sim 5 \times 10^{-24}$ kg-m/s, so this time $\lambda \cong 10^{-10}$ meter. Since this is of the order of the interatomic spacing in the nickel crystal, diffraction effects should be, and are indeed, observed.

QUESTIONS

1. Define or describe briefly what is meant by the following terms: **(a)** blackbody radiation; **(b)** photons; **(c)** matter wave; **(d)** absorption spectrum; and **(e)** photoelectric effect.
2. Why do you suppose the term "blackbody" radiation is used to describe the isothermal radiation emitted by a body in equilibrium at a temperature T?
3. Describe in operational terms the meaning of the following statements: **(a)** The surface of the sun has a temperature of 5800 K. **(b)** Barnard's star has a surface temperature of 3000 K.
4. List the assumptions which Planck introduced to obtain the correct blackbody spectrum? How did his result compare with the classical Rayleigh–Jeans formula?
5. What is the photoelectric effect? What new physical assumptions did Einstein introduce to obtain his basic result in (34-5)?
6. Compare and contrast the assumptions made by Einstein in connection with the photoelectric effect and those made by Planck in connection with his work on blackbody radiation.
7. Can a stream of photons undergo diffraction or interference? Explain.
8. What is Compton scattering? What was the significance of Compton's original experiments in 1923?
9. What is meant by the *emission spectrum* of an element? How is it

related to the absorption spectrum of that element?

10. Describe in physical terms why the emission spectrum of an element will in general appear to exhibit more lines than the corresponding absorption spectrum. What is the nature of the lines missing from the absorption spectrum?
11. Explain in physical terms why the solar spectrum contains only absorption lines. Why might one expect to see emission lines from the sun during a complete solar eclipse?
12. It has been recently discovered that we (and, presumably, the entire universe) are immersed in "three-degree radiation." Explain what you think this means and how this fact could be discovered experimentally. (See Problem 7.)
13. List the basic postulates of the Bohr model and for each one state that feature of classical physics with which it is inconsistent.
14. In what way or ways did the Bohr model fail? Why, besides its historical value, is it still of interest today?
15. Describe an experiment that confirms the existence of matter waves. Explain why matter waves manifest themselves only at the microscopic level.
16. How do you account for the fact that the existence of matter waves remained undetected until relatively recently?

PROBLEMS

1. Calculate the rate of emission per unit area per unit wavelength interval at the wavelength of 6000 Å from a blackbody of temperature: **(a)** 6000 K; **(b)** 1000 K; **(c)** 300 K.
2. By integrating (34-2) over all wavelengths derive the *Stefan–Boltzmann* law

$$\mathcal{F} = \sigma T^4$$

for the total $\mathcal{F}$ rate at which energy is radiated per unit area from a blackbody. Express the Stefan–Boltzmann constant σ, whose value is 5.7×10^{-8} W/m²-K⁴, as a definite integral. (*Hint*: Change dummy variables to remove the T dependence from the integrand for $\mathcal{F}$.)

3. Assuming that you radiate as a blackbody at 310 K, calculate your rate of radiation per unit area by use of the Stefan–Boltzmann law.
4. Assuming that the sun radiates 4.0×10^{26} watts and is a blackbody at 5800 K, calculate its surface area and radius. Compare with the known value of 7.0×10^8 meters for the latter and account for any differences.
5. Show that the intensity I in (34-2) is a maximum at that value for wavelength λ_m which satisfies

$$\lambda_m T = \text{constant}$$

where the constant has the value 0.2898×10^{-2} m-K.

6. By use of the results of Problem 5 find the value λ_m at which I in (34-2) is a maximum for **(a)** the surface of the sun at 5800 K and **(b)** the human body at 310 K. State in each case in what part of the electromagnetic spectrum the peak in the intensity occurs.
7. It has been recently discovered that we are embedded in 2.7-K blackbody radiation (often called the *three-degree radiation*). By use of the result of Problem 5, calculate the value λ_m at which this radiation has its peak. Does the fact that the earth's atmosphere is opaque for wavelengths just below $\cong$0.5 cm affect our ability to measure the full spectrum of this radiation?
8. In a certain experiment, it is found that the lowest wavelength of light

that will produce photoelectrons from a cesium sample is 7.0×10^{-7} meter. Calculate **(a)** the associated value for the threshold frequency ν_c and **(b)** the value of the photoelectric work function for the given sample.

9. The photoelectric work function for a certain sample of copper is 3.0 eV.
   (a) What is ν_c for this sample?
   (b) If light of wavelength 3000 Å is incident on the sample, what is the maximum kinetic energy of the ejected electrons?

10. Monochromatic light of wavelength 6500 Å and of intensity 2.0 W/m^2 is normally incident on a surface. Calculate:
   (a) The energy of the photons in the beam.
   (b) The number of photons incident on the surface per unit area per unit time.

11. Assuming that the total rate of radiation of 4.0×10^{26} watts by the sun is at a wavelength of 5500 Å, calculate the number of photons the sun emits per second.

12. Consider the Compton scattering event in Figure 34-5.
   (a) Show that the conservation-of-energy relation is very generally given by

$$\frac{hc}{\lambda} + mc^2 = \frac{hc}{\lambda'} + c[p^2 + m^2c^2]^{1/2}$$

   where m is the mass of the electron.
   (b) Write down the corresponding relations implied by the law of momentum conservation in terms of λ', p, ϕ, and θ.
   (c) Eliminate the parameters p and ϕ from these relations and thus confirm the validity of (34-6).

13. Suppose that a beam of 2.0×10^5 eV photons is scattered by the electrons in a carbon target.
   (a) What is the wavelength associated with these photons?
   (b) What is the wavelength of those photons scattered through an angle of 90°?
   (c) What is the energy of the scattered photons that emerge at an angle of 60° relative to the incident direction?

14. Calculate numerical values for:
   (a) The electron Compton wavelength.
   (b) The proton Compton wavelength.
   (c) The pion Compton wavelength ($m_\pi = 274m$, with m the electron mass).

15. Calculate the frequency of the line with the longest wavelength in the Balmer series of hydrogen. What is the wavelength of this line?

16. Calculate the frequency and the wavelength of the lines of shortest and longest wavelength in the Lyman series of hydrogen.

17. Prove that the energy levels $E_n(Z)$ of a hydrogenlike atom whose nucleus carries a charge Ze is

$$E_n(Z) = Z^2 E_n$$

with E_n the corresponding energy of ordinary hydrogen. (*Hint*: Repeat the same steps used to derive (34-13) but using the form of the Coulomb force appropriate to this case.)

18. The *binding energy* of an electron in an atom is the minimum energy required to remove it from the atom. Calculate the binding energy of the electron in: **(a)** hydrogen; **(b)** He^+; and **(c)** Li^{++}. (*Hint*: Use the result of Problem 17.)

19. Prove that the speed v_n of an electron in hydrogen in a state corresponding to the quantum number n is

$$v_n = \frac{e^2}{2\epsilon_0 hn}$$

What is the radius of the electron's orbit?

20. Calculate, by use of the result of Problem 19, the time required for an

electron to make a complete orbit when it is in an allowable state.

21. Calculate the de Broglie wavelength associated with a man of mass 80 kg and traveling at a speed of 5 m/s. Would you expect to see diffraction effects in this case? Explain.

22. What is the de Broglie wavelength associated with:
(a) A 2.0 keV electron?
(b) A proton of the same energy?

23. Prove that the Bohr condition in (34-9) is equivalent to the statement that the circumference of the electron orbit in hydrogen is an integral number of de Broglie wavelengths.

APPENDIX A Differentiation formulas

The purpose of this appendix is to list for reference purposes a number of formulas that are useful in calculating derivatives.

Let $x(t)$, $x_1(t)$, and $x_2(t)$ represent functions of the independent variable t, and let α be a constant independent of time and n a positive integer. Then

$$\frac{d}{dt}(\alpha t^n) = n\alpha t^{n-1} \tag{A-1}$$

$$\frac{d}{dt}[x_1(t) + \alpha x_2(t)] = \frac{dx_1}{dt} + \alpha \frac{dx_2}{dt} \tag{A-2}$$

$$\frac{d}{dt}[x_1(t)x_2(t)] = x_1 \frac{dx_2}{dt} + x_2 \frac{dx_1}{dt} \tag{A-3}$$

$$\frac{d}{dt} f(x(t)) = \frac{df}{dx}\frac{dx}{dt} \tag{A-4}$$

where, in (A-4), f is a function of the independent variable t only through its explicit dependency on the function $x(t)$. The relation in (A-4) is known as the *chain rule.*

To give some idea as to how these results are obtained, let us derive (A-1). Applying the definition of a derivative

$$\frac{dx}{dt} = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}$$

to the function $x = \alpha t^n$, we obtain

$$\frac{d}{dt}(\alpha t^n) = \lim_{\Delta t \to 0} \frac{\alpha(t+\Delta t)^n - \alpha t^n}{\Delta t}$$

By use of the binomial expansion

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \cdots + b^n$$

and the fact that n is a positive integer, this may be simplified to the desired result:

$$\frac{d(\alpha t^n)}{dt} = \lim_{\Delta t \to 0} \alpha \frac{t^n + nt^{n-1}\Delta t + \cdots + (\Delta t)^n - t^n}{\Delta t}$$
$$= \alpha n t^{n-1}$$

More generally, it can be shown that (A-1) is valid for n any positive or negative number.

The validity of (A-2), (A-3), and (A-4) and related formulas may be established in a similar way.

APPENDIX B Planetary data

Planet	*Radius* (10^3 km)	*Radius* (earth = 1)	*Distance from sun* (10^6 km)	*Mass* (5.98×10^{24} kg)	*Density* (10^3 kg/m^3)	*Rotation period* (sideral)	*Period* (years)
Mercury	2.4	0.38	58	0.06	5.4	59^d	0.24
Venus	6.1	0.95	108	0.82	5.3	243^d	0.62
Earth	6.37	1.00	150	1.00	5.5	$23^h56^m04^s$	1.00
Mars	3.39	0.53	228	0.11	3.94	$24^h37^m23^s$	1.88
Jupiter	70	11.0	778	318	1.33	9^h55^m	11.9
Saturn	58	9.1	1426	95	0.70	10^h38^m	29.5
Uranus	24	3.7	2869	14.5	1.56	10^h49^m	84
Neptune	25	3.5	4495	17.2	2.27	16^h(?)	165
Pluto	6.0	0.46(?)	5900	0.06(?)	?	6.4^d(?)	248

The sun: Mass, 330,000 $M_\oplus = 1.99 \times 10^{30}$ kg; mean radius, 6.95×10^5 km; mean density, 1.42×10^3 kg/m^3

The moon: Mass, .0123 $M_\oplus$; mean radius, 1.74×10^3 km; mean density, 3.36×10^3 kg/m^3; mean distance from earth, 3.8×10^5 km

The earth: $M_\oplus = 5.98 \times 10^{24}$ kg; mean radius, 6.35×10^3 km; mean density, 5.5×10^3 kg/m^3

APPENDIX C The derivatives of the sine and the cosine

The purpose of this appendix is to derive the rules for differentiating the sine and the cosine functions. To this end, let us first establish the limit

$$\lim_{\theta \to 0} \frac{\sin \theta}{\theta} = 1 \tag{C-1}$$

where the angle θ is expressed in radians. The fact that for sufficiently small angles, $\theta \cong \sin \theta$ is easily established by reference to a table of trigonometric functions. Expressing θ in radians we find the values $\sin 0.10 = 0.09983$, $\sin 0.05 = 0.04998$, $\sin 0.01 = 0.01000$, and, more generally, for $\theta \leqslant 0.10$ rad ($\cong 6°$) we find that θ and $\sin \theta$ differ from each other by less than 1 percent.

To establish the validity of (C-1), consider in Figure C-1 a circle of radius R centered at A, and let AB and AC be two radii subtending an angle θ. Construct from the point B the line BE tangent to the circle at B and the line BD to be perpendicular to AC. The angle between BD and BE is also θ. By

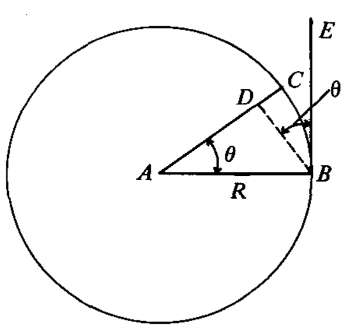


Figure C-1

definition of a radian, the length of the arc BC is $R\theta$ and thus

$$\theta = \frac{\text{arc } BC}{R}$$

Further, in the right triangle ABD,

$$\sin\theta = \frac{BD}{R}$$

and thus by division we find that

$$\frac{\sin\theta}{\theta} = \frac{BD}{\text{arc } BC}$$

In the limit as $\theta \to 0$, since BD becomes parallel to BE, it follows that the length of BD must become more and more equal to the arc BC. The validity of (C-1) is thus established.

We shall now use (C-1) to calculate the derivative of $\sin\theta$. According to the definition of a derivative, we have

$$\begin{aligned}\frac{d}{d\theta}\sin\theta &= \lim_{\Delta\theta\to 0}\frac{\sin(\theta+\Delta\theta)-\sin\theta}{\Delta\theta}\\ &= \lim_{\Delta\theta\to 0}\frac{\sin\theta\cos\Delta\theta+\sin\Delta\theta\cos\theta-\sin\theta}{\Delta\theta}\end{aligned} \tag{C-2}$$

where the second equality follows by use of the identity

$$\sin(A+B) = \sin A\cos B + \cos A \sin B$$

Since $\cos 0 = 1$, it follows that for small $\Delta\theta$, $\cos\Delta\theta \cong 1$. Hence (C-2) becomes

$$\begin{aligned}\frac{d}{d\theta}\sin\theta &= \lim_{\Delta\theta\to 0}\frac{\cos\theta\sin\Delta\theta}{\Delta\theta}\\ &= \cos\theta\end{aligned} \tag{C-3}$$

where the second equality follows by use of (C-1).

Similarly, the derivative of the cosine is found to be

$$\begin{aligned}\frac{d}{d\theta}\cos\theta &= \lim_{\Delta\theta\to 0}\frac{\cos(\theta+\Delta\theta)-\cos\theta}{\Delta\theta}\\ &= \lim_{\Delta\theta\to 0}\frac{\cos\theta\cos\Delta\theta-\sin\theta\sin\Delta\theta-\cos\theta}{\Delta\theta}\\ &= \lim_{\Delta\theta\to 0}\frac{-\sin\theta\sin\Delta\theta}{\Delta\theta}\\ &= -\sin\theta\end{aligned} \tag{C-4}$$

where the second equality follows from the identity

$$\cos(A+B) = \cos A\cos B - \sin A\sin B$$

and the fourth from (C-1).

Applying the chain rule (A-4), and (C-3), (C-4), we find that if h is a function of t, then

$$\frac{d}{dt}\sin[h(t)] = \cos[h(t)]\frac{dh}{dt} \tag{C-5}$$

$$\frac{d}{dt}\cos[h(t)] = -\sin[h(t)]\frac{dh}{dt} \tag{C-6}$$

For the special case, $h(t) = \omega t + \alpha$, with ω and α constants, these formulas reduce to (6-17) and (6-18), respectively.

APPENDIX D Proof of Theorem I

To prove Theorem I in (11-13), we start with the equations of motion for the N constituents of the system:

$$
\begin{aligned}
m_1 \frac{d\mathbf{v}_1}{dt} &= \mathbf{F}_1 + \mathbf{F}_{12} + \mathbf{F}_{13} + \cdots \\
m_2 \frac{d\mathbf{v}_2}{dt} &= \mathbf{F}_2 + \mathbf{F}_{21} + \mathbf{F}_{23} + \cdots \\
\vdots \quad & \quad \vdots \quad \vdots \quad \vdots \\
m_N \frac{d\mathbf{v}_N}{dt} &= \mathbf{F}_N + \mathbf{F}_{N1} + \mathbf{F}_{N2} + \cdots
\end{aligned}
\tag{D-1}
$$

where $\mathbf{F}_i$ is the *external* force acting on the ith particle and $\mathbf{F}_{ij}$ is the force on the ith particle produced by the jth. Let us add together these equations. On the left-hand side (LHS) we obtain

$$
\begin{aligned}
\text{LHS} &= m_1 \frac{d\mathbf{v}_1}{dt} + m_2 \frac{d\mathbf{v}_2}{dt} + \cdots + m_N \frac{d\mathbf{v}_N}{dt} \\
&= \frac{d}{dt}(m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \cdots + m_N\mathbf{v}_N) \\
&= M\frac{d\mathbf{V}_c}{dt}
\end{aligned}
\tag{D-2}
$$

where the third equality follows by use of (11-14). On the right-hand side (RHS) we obtain by contrast two types of terms: (1) those due to the external forces; and (2) those due to the internal forces. According to the third law,

these latter terms cancel in pairs. For example, the term $\mathbf{F}_{12}$ from the first of (D-1) cancels with the $\mathbf{F}_{21}$ term from the second. More generally, the term $\mathbf{F}_{ij}$ from the ith equation cancels with the term $\mathbf{F}_{ji}$ from the jth. Accordingly, we obtain

$$\text{RHS} = \mathbf{F}_1 + \mathbf{F}_2 + \cdots + \mathbf{F}_N \tag{D-3}$$

and, combining (D-2) and (D-3), the result in (11-13) follows.

APPENDIX E Proof of the parallel-axis theorem

To establish the *parallel-axis theorem*, consider the situation in Figure 12-5. Let I be the moment of inertia of the body about the z-axis and I_c the corresponding moment of inertia about a parallel axis through the center of mass at (X_c, Y_c, Z_c). The perpendicular distance D between these axes satisfies

$$D^2 = X_c^2 + Y_c^2 \tag{E-1}$$

where

$$X_c = \frac{1}{M}\sum m_i x_i \qquad Y_c = \frac{1}{M}\sum m_i y_i \tag{E-2}$$

and where $M = \Sigma\, m_i$, with m_i the mass of the particle located at the point (x_i, y_i, z_i).

In terms of this notation,

$$I = \sum m_i (x_1^2 + y_i^2)$$

and

$$I_c = \sum m_i [(X_c - x_i)^2 + (Y_c - y_i)^2] \tag{E-3}$$

since the quantity in the square bracket is the square of the perpendicular distance from m_i to the axis through the mass center. On expanding the square bracket, we obtain

$$\begin{aligned}
I_c &= \sum m_i [x_i^2 + y_i^2 + X_c^2 + Y_c^2 - 2X_c x_i - 2Y_c y_i] \\
&= \sum m_i (x_i^2 + y_i^2) + (X_c^2 + Y_c^2) \sum m_i - 2X_c \sum m_i x_i - 2Y_c \sum m_i y_i \\
&= I + (X_c^2 + Y_c^2)M - 2X_c M X_c - 2Y_c M Y_c \\
&= I - M(X_c^2 + Y_c^2) \\
&= I - MD^2
\end{aligned}$$

The third equality follows by use of (12-5) and the last by (E-1). Finally, transposing we obtain the parallel-axis theorem in (12-7).

APPENDIX F Derivation of the wave equation

To derive (18-1) for small-displacement waves on a string, consider in Figure 18-8 an infinitesimal element of the string of length Δx at an instant when it has the displacement y. If μ is the mass per unit length of the string, the mass of this element is $\mu\ \Delta x$. Applying Newton's second law, we may write

$$F_y = \mu\ \Delta x \frac{\partial^2 y}{\partial t^2} \tag{F-1}$$

where F_y is the vertical component of the force acting on the mass element and $\partial^2 y/\partial t^2$ is its acceleration.

To express F_y in terms of the tension T_0 in the string, let us assume the displacements to be small so that the angle θ satisfies

$$\tan\theta \cong \sin\theta \cong \theta \tag{F-2}$$

Reference to Figure 18-8 shows that

$$F_y = T_0 \sin(\theta + \Delta\theta) - T_0 \sin\theta \cong T_0(\theta + \Delta\theta) - T\theta = T_0\Delta\theta$$

Moreover, since $\tan\theta = \partial y/\partial x$, we find by use of (F-2) that

$$\Delta\theta = \Delta\left(\frac{\partial y}{\partial x}\right) = \frac{\partial^2 y}{\partial x^2}\Delta x$$

where the second equality follows provided that Δx is sufficiently small. Hence F_y becomes

$$F_y = T_0\,\Delta\theta = T_0\,\Delta x \frac{\partial^2 y}{\partial x^2} \tag{F-3}$$

Substituting into (F-1) and canceling the common factor Δx we obtain the wave equation in (18-1).

APPENDIX G Table of trigonometric functions

Degrees	*Radians*	*Sine*	*Cosine*	*Tangent*	*Cotangent*		
0	0	0	1.0000	0	——	1.5708	90
1	.0175	.0175	.9998	.0175	57.290	1.5533	89
2	.0349	.0349	.9994	.0349	28.636	1.5359	88
3	.0524	.0523	.9986	.0524	19.081	1.5184	87
4	.0698	.0698	.9976	.0699	14.301	1.5010	86
5	.0873	.0872	.9962	.0875	11.430	1.4835	85
6	.1047	.1045	.9945	.1051	9.5144	1.4661	84
7	.1222	.1219	.9925	.1228	8.1443	1.4486	83
8	.1396	.1392	.9903	.1405	7.1154	1.4312	82
9	.1571	.1564	.9877	.1584	6.3138	1.4137	81
10	.1745	.1736	.9848	.1763	5.6713	1.3963	80
11	.1920	.1908	.9816	.1944	5.1446	1.3788	79
12	.2094	.2079	.9781	.2126	4.7046	1.3614	78
13	.2269	.2250	.9744	.2309	4.3315	1.3439	77
14	.2443	.2419	.9703	.2493	4.0108	1.3265	76
		Cosine	*Sine*	*Cotangent*	*Tangent*	*Radians*	*Degrees*

Degrees	*Radians*	*Sine*	*Cosine*	*Tangent*	*Cotangent*		
15	.2618	.2588	.9659	.2679	3.7321	1.3090	75
16	.2793	.2756	.9613	.2867	3.4874	1.2915	74
17	.2967	.2924	.9563	.3057	3.2709	1.2741	73
18	.3142	.3090	.9511	.3249	3.0777	1.2566	72
19	.3316	.3256	.9455	.3443	2.9042	1.2392	71
20	.3491	.3420	.9397	.3640	2.7475	1.2217	70
21	.3665	.3584	.9336	.3839	2.6051	1.2043	69
22	.3840	.3746	.9272	.4040	2.4751	1.1868	68
23	.4014	.3907	.9205	.4245	2.3559	1.1694	67
24	.4189	.4067	.9135	.4452	2.2460	1.1519	66
25	.4363	.4226	.9063	.4663	2.1445	1.1345	65
26	.4538	.4384	.8988	.4877	2.0503	1.1170	64
27	.4712	.4540	.8910	.5095	1.9626	1.0996	63
28	.4887	.4695	.8829	.5317	1.8807	1.0821	62
29	.5061	.4848	.8746	.5543	1.8040	1.0647	61
30	.5236	.5000	.8660	.5774	1.7321	1.0472	60
31	.5411	.5150	.8572	.6009	1.6643	1.0297	59
32	.5585	.5299	.8480	.6249	1.6003	1.0123	58
33	.5760	.5446	.8387	.6494	1.5399	.9948	57
34	.5934	.5592	.8290	.6745	1.4826	.9774	56
35	.6109	.5736	.8192	.7002	1.4281	.9599	55
36	.6283	.5878	.8090	.7265	1.3764	.9425	54
37	.6458	.6018	.7986	.7536	1.3270	.9250	53
38	.6632	.6157	.7880	.7813	1.2799	.9076	52
39	.6807	.6293	.7771	.8098	1.2349	.8901	51
40	.6981	.6428	.7660	.8391	1.1918	.8727	50
41	.7156	.6561	.7547	.8693	1.1504	.8552	49
42	.7330	.6691	.7431	.9004	1.1106	.8378	48
43	.7505	.6820	.7314	.9325	1.0724	.8203	47
44	.7679	.6947	.7193	.9657	1.0355	.8029	46
45	.7854	.7071	.7071	1.0000	1.0000	.7854	45
		Cosine	*Sine*	*Cotangent*	*Tangent*	*Radians*	*Degrees*

APPENDIX H Table of e^x and e^{-x}

x	e^x	e^{-x}	x	e^x	e^{-x}
0.0	1.0000	1.0000	3.0	20.086	.04979
0.1	1.1052	.90484	3.1	22.198	.04505
0.2	1.2214	.81873	3.2	24.533	.04076
0.3	1.3499	.74082	3.3	27.113	.03688
0.4	1.4918	.67032	3.4	29.964	.03337
0.5	1.6487	.60653	3.5	33.115	.03020
0.6	1.8221	.54881	3.6	36.598	.02732
0.7	2.0138	.49659	3.7	40.447	.02472
0.8	2.2255	.44933	3.8	44.701	.02237
0.9	2.4596	.40657	3.9	49.402	.02024
1.0	2.7183	.36788	4.0	54.598	.01832
1.1	3.0042	.33287	4.1	60.340	.01657
1.2	3.3201	.30119	4.2	66.686	.01500
1.3	3.6693	.27253	4.3	73.700	.01357
1.4	4.0552	.24660	4.4	81.451	.01228
1.5	4.4817	.22313	4.5	90.017	.01111
1.6	4.9530	.20190	4.6	99.484	.01005
1.7	5.4739	.18268	4.7	109.95	.00910
1.8	6.0496	.16530	4.8	121.51	.00823
1.9	6.6859	.14957	4.9	134.29	.00745
2.0	7.3891	.13534	5.0	148.41	.00674
2.1	8.1662	.12246	5.5	244.69	.00409
2.2	9.0250	.11080	6.0	403.43	.00248
2.3	9.9742	.10026	6.5	665.14	.00150
2.4	11.023	.09072	7.0	1096.6	.00091
2.5	12.182	.08208	7.5	1808.0	.00055
2.6	13.464	.07427	8.0	2981.0	.00034
2.7	14.880	.06721	8.5	4914.8	.00020
2.8	16.445	.06081	9.0	8103.1	.00012
2.9	18.174	.05502	9.5	13360	.00007
3.0	20.086	.04979	10.0	22026	.00005

APPENDIX J Experimental values of basic physical constants[a]

Speed of light	c	$2.997924562(11) \times 10^{8}$ m/s
Gravitational constant	G	$6.6732(31) \times 10^{-11}$ N-m²/kg²
Avogadro's number	N_0	$6.022169(40) \times 10^{23}$ atoms/mole
Boltzmann's constant	k	$1.380622(59) \times 10^{-23}$ J/K
Gas constant	R	8.31434(35) J/mole-K
Triple point of water	T_t	273.16 K (exact)
Electron mass	m_e	$9.109558(54) \times 10^{-31}$ kg
Proton mass	m_p	$1.672614(11) \times 10^{-27}$ kg
Quantum of electric charge	e	$1.6021917(70) \times 10^{-19}$ C
Permittivity of free space	ϵ_0	$8.8541853(59) \times 10^{-12}$ F/m
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ H/m (exact)
Planck's constant	h	$6.626196(50) \times 10^{-34}$ J-s
Bohr radius	a_0	$5.2917715(81) \times 10^{-11}$ m
Electron Compton wavelength	$\dfrac{h}{m_e c}$	$2.4263096(74) \times 10^{-12}$ m

[a]The values in this table (except for c) are adapted from, B. N. Taylor, W. H. Parker, and D. N. Langenberg, *Fundamental Constants and Quantum Electrodynamics*, New York, Academic Press, 1969. The value for c is taken from the article by K. M. Evenson *et al.*, *Phys. Rev. Lett.* **29**, 1346 (1972). The numbers in parentheses refer to the error in the last digits of the listed values.

APPENDIX K Units for physical quantities

Quantity	Symbol	Derived Unit	SI Unit
Acceleration	$\mathbf{a}$	m/s^2	m/s^2
Angle	θ	radian (rad)	—
Angular acceleration	α	rad/s^2	s^{-2}
Angular momentum	$\mathbf{L}$	J-s	$kg\text{-}m^2/s$
Angular velocity	ω	rad/s	s^{-1}
Area	S, A	m^2	m^2
Capacitance	C	farad(F)	$A^2\text{-}s^4/kg\text{-}m^2$
Charge	q, e	coulomb(C)	A-s
Charge density			
Volume	ρ	C/m^3	$A\text{-}s/m^3$
Surface	σ	C/m^2	$A\text{-}s/m^2$
Linear	λ	C/m	A-s/m
Coefficient of linear expansion	α	K^{-1}	K^{-1}
Conductivity	σ	$(\Omega\text{-}m)^{-1}$	$A^2\text{-}s^3/kg\text{-}m^3$
Current	i, I	ampere(A)	A
Current density	$\mathbf{J}$	A/m^2	A/m^2
Electric dipole moment	$\mathbf{p}$	C-m	A-s-m
Electric displacement	$\mathbf{D}$	C/m^2	$A\text{-}s/m^2$
Electric field	$\mathbf{E}$	V/m	$kg\text{-}m/A\text{-}s^3$
Electric flux	Φ, Φ_E	V-m	$kg\text{-}m^3/A\text{-}s^3$
Electromotive force	$\mathscr{E}$	volt(V)	$kg\text{-}m^2/A\text{-}s^3$
Energy	E, T, V	joule(J)	$kg\text{-}m^2/s^2$

Quantity	*Symbol*	*Derived Unit*	*SI Unit*
Entropy	S	J/K	kg-m^2/s^2-K
Flow rate	V_R	m^3/s	m^3/s
Force	$\mathbf{F}$	newton(N)	kg-m/s^2
Frequency	ν	hertz(Hz)	s^{-1}
Gravitational field strength	$\mathbf{g}$	m/s^2	m/s^2
Heat	Q	J	kg-m^2/s^2
Inductance	L	henry(H)	kg-m^2/A^2-s^2
Length	l, d, L	m	m
Magnetization	$\mathbf{M}$	A/m	A/m
Magnetic dipole moment	$\boldsymbol{\mu}$	N-m/T	A-m^2
Magnetic field	$\mathbf{H}$	A/m	A/m
Magnetic flux	Φ_B	weber(Wb)	kg-m^2/A-s^2
Magnetic induction	$\mathbf{B}$	tesla (T = Wb/m^2)	kg/A-s^2
Mass	m, M	kg	kg
Mass density	ρ	kg/m^3	kg/m^3
Molar specific heat	C	J/mole-K	kg-m^2/s^2-mole-K
Moment of inertia	I	kg-m^2	kg-m^2
Momentum	$\mathbf{p}$	N-s	kg-m/s
Particle density	n	molecules/m^3	m^{-3}
Permeability of free space	μ_0	H/m	kg-m/A^2-s^2
Permittivity of free space	ϵ_0	F/m	A^2-s^4/kg-m^3
Polarization	$\mathbf{P}$	C/m^2	A-s/m^2
Potential	V	volt(V)	kg-m^2/A-s^3
Power	P	watt(W)	kg-m^2/s^3
Poynting vector	$\mathbf{N}$	W/m^2	kg/s^3
Pressure	P	N/m^2	kg/m-s^2
Resistance	R	ohm(Ω)	kg-m^2/A^2-s^3
Resistivity	ρ	Ω-m	kg-m^3/A^2-s^3
Specific heat	c	J/kg-K	m^2/s^2-K
Spring constant	k	N/m	kg/s^2
Temperature	T	kelvin(K)	K
Time	t	second	s
Torque	$\boldsymbol{\tau}$	N-m	kg-m^2/s^2
Velocity (speed)	$\mathbf{v}$	m/s	m/s
Volume	V	m^3	m^3
Wave number	k	m^{-1}	m^{-1}
Wavelength	λ	m	m
Work	W	J	kg-m^2/s^2

APPENDIX L Conversion to the Gaussian system of units

The form of the equations of electromagnetism used in this text is applicable provided all quantities are expressed in SI (mks) units. To convert any of these equations to the corresponding Gaussian form, replace each of the SI symbols for the quantities listed below under "SI" by the corresponding one listed under "Gaussian." For example, under this replacement the relation $C = \varepsilon_0 A/d$ becomes $4\pi\varepsilon_0 C = \varepsilon_0 A/d$, so that the corresponding Gaussian form is $C = A/4\pi d$. Note that in this process the symbols for mass, length, time, and other mechanical quantities are not changed and that any noncancelling factors of $(\varepsilon_0\mu_0)^{-\frac{1}{2}}$ are to be replaced by the speed of light c.

Physical quantity	*SI*	*Gaussian*
Capacitance	C	$4\pi\varepsilon_0 C$
Conductivity	σ	$4\pi\varepsilon_0\sigma$
Charge (as well as charge density, current, current density and polarization)	$q[\rho, i, \mathbf{J}, \mathbf{P}]$	$\sqrt{4\pi\varepsilon_0}\, q[\rho, i, \mathbf{J}, \mathbf{P}]$
Dielectric constant	κ	$\varepsilon_0\kappa$
Displacement vector	$\mathbf{D}$	$\sqrt{\varepsilon_0/4\pi}\, \mathbf{D}$
Electric field (and potential and flux)	$\mathbf{E}[V, \Phi]$	$(4\pi\varepsilon_0)^{-\frac{1}{2}}\, \mathbf{E}[V, \Phi]$
Inductance	L	$L/4\pi\varepsilon_0$
Magnetic field	$\mathbf{H}$	$(4\pi\mu_0)^{-\frac{1}{2}}\, \mathbf{H}$
Magnetic induction (and flux)	$\mathbf{B}[\Phi_m]$	$\sqrt{\mu_0/4\pi}\, \mathbf{B}[\Phi_m]$
Magnetization	$\mathbf{M}$	$\sqrt{4\pi/\mu_0}\, \mathbf{M}$
Relative permeability	κ_m	$\mu_0\kappa_m$
Resistance	R	$R/4\pi\varepsilon_0$

The table below lists side by side equivalent amounts of various physical quantities in SI and Gaussian units. The symbol α represents the ratio $c/10^8$m/s, which, according to (30-3), has the value 2.997925. For many purposes the approximation $\alpha = 3$ is sufficient.

Physical quantity	*Symbol*	*SI*	*Gaussian*
Capacitance	C	1 F	$\alpha^2 \times 10^{11}$cm
Charge	q	1 C	$\alpha \times 10^9$ statcoul
Charge density	ρ	1 C/m^3	$\alpha \times 10^3$ statcoul/cm^3
Conductivity	σ	1 $(\Omega\text{-m})^{-1}$	$\alpha^2 \times 10^9$ s^{-1}
Current	i	1 A	$\alpha \times 10^9$ statamps
Current density	J	1 A/m^2	$\alpha \times 10^5$ statamp/cm^2
Displacement vector	D	1 C/m^2	$4\pi\alpha \times 10^5$ statvolt/cm
Electric field	E	1 V/m	$\alpha^{-1} \times 10^{-4}$ statvolt/cm
Energy (work)	$U(W)$	1 J	10^7 erg
Force	F	1 N	10^5 dynes
Length	l, d	1 m	10^2 cm
Magnetic field	H	1 A/m	$4\pi \times 10^{-3}$ oersted
Magnetic flux	Φ_m	1 Wb	10^8 maxwells
Magnetic induction	B	1 T	10^4 gauss
Magnetization	M	1 A/m	$(4\pi)^{-1} \times 10^{-3}$ oersted
Mass	m	1 kg	10^3g
Polarization	P	1 C/m^2	$\alpha \times 10^5$ statcoul/cm^2
Potential	V	1 V	$\alpha^{-1} \times 10^{-2}$ statvolt
Power	P	1 W	10^7 erg/s
Resistance	R	1 Ω	$\alpha^{-2} \times 10^{-11}$ s/cm
Time	t	1 s	1 s

THE PERIODIC TABLE OF THE ELEMENTS

The number directly above the chemical symbol for each element is its atomic number Z, and the number below is its atomic weight in amu. For unstable elements, the atomic weight of the most stable isotope is given in parentheses.

1a	2a	3b	4b	5b	6b	7b	8			1b	2b	3a	4a	5a	6a	7a	0
1 **H** 1.01																	2 **He** 4.00
3 **Li** 6.94	4 **Be** 9.01											5 **B** 10.81	6 **C** 12.01	7 **N** 14.01	8 **O** 16.00	9 **F** 19.00	10 **Ne** 20.18
11 **Na** 22.99	12 **Mg** 24.31											13 **Al** 26.98	14 **Si** 28.09	15 **P** 30.97	16 **S** 32.06	17 **Cl** 35.45	18 **Ar** 39.95
19 **K** 39.10	20 **Ca** 40.08	21 **Sc** 44.96	22 **Ti** 47.90	23 **V** 50.94	24 **Cr** 52.00	25 **Mn** 54.94	26 **Fe** 55.85	27 **Co** 58.93	28 **Ni** 58.71	29 **Cu** 63.55	30 **Zn** 65.37	31 **Ga** 69.72	32 **Ge** 72.59	33 **As** 74.92	34 **Se** 78.96	35 **Br** 79.90	36 **Kr** 83.80
37 **Rb** 85.47	38 **Sr** 87.62	39 **Y** 88.91	40 **Zr** 91.22	41 **Nb** 92.91	42 **Mo** 95.94	43 **Tc** (99)	44 **Ru** 101.07	45 **Rh** 102.91	46 **Pd** 106.4	47 **Ag** 107.87	48 **Cd** 112.	49 **In** 114.82	50 **Sn** 118.69	51 **Sb** 121.75	52 **Te** 127.60	53 **I** 126.90	54 **Xe** 131.30
55 **Cs** 132.91	56 **Ba** 137.34	57-71 *	72 **Hf** 178.49	73 **Ta** 180.95	74 **W** 183.85	75 **Re** 186.2	76 **Os** 190.2	77 **Ir** 192.2	78 **Pt** 195.09	79 **Au** 196.97	80 **Hg** 200.59	81 **Tl** 204.37	82 **Pb** 207.19	83 **Bi** 208.98	84 **Po** (210)	85 **At** (210)	86 **Rn** (222)
87 **Fr** (223)	88 **Ra** (226)	89-103 **	104 **Ku** (260)														

* Lanthanide Series	57 **La** 138.91	58 **Ce** 140.12	59 **Pr** 140.91	60 **Nd** 144.24	61 **Pm** (147)	62 **Sm** 150.35	63 **Eu** 151.96	64 **Gd** 157.25	65 **Tb** 158.92	66 **Dy** 162.50	67 **Ho** 164.93	68 **Er** 167.26	69 **Tm** 168.93	70 **Yb** 173.04	71 **Lu** 174.97
** Actinide Series	89 **Ac** (227)	90 **Th** 232.04	91 **Pa** (231)	92 **U** 238.03	93 **Np** (237)	94 **Pu** (242)	95 **Am** (243)	96 **Cm** (247)	97 **Bk** (247)	98 **Cf** (251)	99 **Es** (254)	100 **Fm** (253)	101 **Md** (256)	102 **No** (254)	103 **Lw** (257)

Solutions to Odd-Numbered Problems

Chapter 1

1. 3.15×10^7 s
3. 56 mi/hr
5. 6.0×10^{24} kg
7. 10 s; no
9. 3.1×10^{24} atoms
11. —
13. $\sim 3.1 \times 10^{-10}$ m
15. **(a)** 1.8×10^{-27} g; **(b)** converted to energy
17. **(b)** 1.1×10^3 electron masses

Chapter 2

1. **(a)** 50 km/hr; 67 km/hr; **(b)** 57 km/hr
3. **(a)** 3.0 m/s; **(b)** 2.1 m/s; 5.0 m/s
5. 1.5 yr
7. **(a)** 0.67 m/s^2; **(b)** 1.3 m/s^2
9. **(a)** 8.0 km/hr; **(b)** 40 km/hr^2; **(c)** -8.0 km/hr^2 (deceleration); **(d)** 16 km/hr^2
11. $(6t + 2 + 3\Delta t)$; 11 m/s, 8.3 m/s, 8.00003 m/s; 8.0 m/s
13. **(a)** $(4\alpha t^3 - 3\beta t^2)$; **(b)** $(-at + v_0)$; **(c)** $-6t$; **(d)** $4\alpha t^3$; **(e)** $(2\alpha t^3 - \beta t^2 - 2at + v_0)$
15. **(a)** 3 m/s; **(b)** -2 m/s; **(c)** 2 m/s
17. **(a)** -34 m/s^2; **(b)** 0; **(c)** 44 m/s^2; **(d)** $-24t$; **(e)** $(252t^5 - 48t^2)$
19. **(a)** $-t^2$; **(b)** $(-4.9t^2 + 2)$; **(c)** $(1 + t - t^3/2)$; **(d)** $\left(2 + 2t + \frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4\right)$
21. **(a)** 4.0×10^{-6} s; **(b)** 4.0×10^{-3} m; **(c)** 8.0×10^{-6} s; -2000 m/s
23. 4.1 s; -20 m/s
25. 2.3 s; 23 m/s
27. **(a)** 5.1×10^2 m; **(b)** 20 s; **(c)** 71 m/s
29. —
31. **(a)** 4.0 m/s^2; **(b)** 13 m
33. —

35. (a) 4.3 s; (b) 14 m/s; (c) 44 m/s; 46 m/s
37. (a) $[(-bt^4/12)+3t]$; (b) $[(-bt^4/12)+(3+b/12)t]$; (c) $b/12$

Chapter 3

1. —
3. (b) no
5. —
7. 74°S of E
9. 3.6 m/s^2; 74°; 134°
11. (a) $(5\mathbf{i}-\mathbf{j})$; (b) 7, 2; (c) −3, −4
13. —
15. (a) $\sqrt{13}$; (b) 34°; (c) $(3\mathbf{i}+2\mathbf{j})/\sqrt{13}$
17. —
19. $[(3/2)\mathbf{i}t^2-(1/3)\mathbf{j}t^3]$
21. —
23. $4\pi^2R\cos\lambda/P^2$
25. 0.79 m/s^2
27. (a) 10 m; (b) 1.4 s
29. —
31. 42 m/s
33. $[2R_0\cos^2(\beta+\gamma)/\cos\beta] \times [\tan(\beta+\gamma)-\tan\beta]$; $(90°-\beta)/2$
35. (a) 12 m; (b) 43 m; (c) 21 m/s at an angle of 48° to the horizontal (downward)
37. (a) $[\mathbf{i}(t^2+2t)+3\mathbf{j}t^3]$; (b) $[\mathbf{i}(t^2-3t)+3\mathbf{j}t^3]$
39. (a) $x = ut$; $y = -\frac{1}{2}gt^2+tv_0\sin\alpha-h$
41. $x'(t') = t'(v_0\cos\theta-u)/[1-(uv_0/c^2)\cos\theta]$; $y'(t') = t'v_0\sin\theta/\gamma[1-(uv_0/c^2)\cos\theta]$

Chapter 4

1. (a) 0.50 m/s^2; (b) 0.30 m/s^2
3. 4.0 N
5. 1.0 N
7. (a) 3.6×10^3 N; (b) 3.6×10^3 N; 4.1
9. **(a) 1.54 m/s^2; (b) 18 s; (c) 2.25 $\times$** 10^3 N; tires
11. **(a) 2.2 N; θ = 207.** (b) 6.1 N; $\theta = 189°$
13. (a) 5.0×10^{-2} m/s^2; (b) 2.5×10^{-4} N
15. (a) 1.01×10^3 m/s; (b) 2.7×10^{-3} m/s^2; (c) 2.0×10^{20} N
17. (a) 1.1 N; (b) 0.74 m/s
19. (a) 3.4×10^{-2} m/s^2; (b) 3.4 N; (c) lighter, 3.4 N
21. 3.6×10^{22} N
23. 5.1×10^{24} kg
25. 4.1×10^{16} N
27. **$m_1F/(m_1+m_2)$ = 0.75 N**
29. (a) 18 N; (b) 14 N; (c) 14 N; (d) 8.0 N
31. (a) 1.8 N; (b) 0.60 N for string at A; 1.2 N for string at B
33. (c) 20 N in the right string; 13 N in the left string
35. (a) zero; (b) 15 N in left; 9.8 N in right string below point of stone's attachment
37. 0.65 N; 3.3 m/s^2
39. (a) 2.9×10^3 N; (b) 2.9×10^3 N
41. (a) 4.9 N; (b) 5.1 N; (c) 4.8 N

Chapter 5

1. 1.6 N; directed 68° above $\mathbf{F}_2$
3. 90° between $\mathbf{F}_1$ and $\mathbf{F}_2$; 37° between $\mathbf{F}_2$ and $\mathbf{F}_3$; and 53° between $\mathbf{F}_1$ and $\mathbf{F}_3$
5. w in a; $w/\sqrt{3}$ in b; $2w/\sqrt{3}$ in c
7. (a) 19.6 N (vertically up); (b) 1.7
9. —
11. (a) 2.4 m/s; (b) 9.8×10^{-3}
13. (a) 36.4 m/s; (b) 34.5 m/s
15. (a) 7.3 m/s^2; (b) 1.8 s; (c) 15 m/s
17. v_0^2/RD^2, $-v_0^2\mu_k/RD^2$ with $D = 1+\mu_kv_0t/R$
19. **(a) 9.8 $\times$ 10^5 m; (b) 1.4 $\times$ 10^7 m**
21. —
23. (a) 1.1×10^4 m/s; (b) 2.4×10^3 m/s; (c) 6.2×10^5 m/s
25. (a) 63 rad/s; (b) 19 m/s; (c) 1.2×10^3 m/s^2
27. (a) a_0; (b) a_0/R; (c) $m[a_0^2+(a_0t-v_0)^4/R^2]^{1/2}$
29. 9°
31. (a) 7.8 m/s; (b) 0.55 s
33. —
35. (a) 140 m/s; (b) 4.9×10^{-2} kg/s
37. —

Chapter 6

1. (a) 0.4 s; (b) 2.5 Hz
3. 1.3×10^2 rad/s; 5.0×10^{-2} s
5. (a) 0.5 m; (b) 5.5 m/s; (c) zero; (d) 60 m/s^2
7. 0.15 m; −0.41 m/s; −0.37 m/s^2
9. 0.4 cm
11. 6.3 s
13. (a) 4.9 N/m; (b) zero; (c) $x(t)=0.20\cos[2\pi t/1.8\text{ s}]$ (m)
15. $x(t)=0.19\sin[\pi t/2.0\text{ s}]$ (m) $v(t)=0.30\cos[\pi t/2.0\text{ s}]$ (m/s)
17. 66 cm
19. —
21. (a) 12 cm, zero; (b) 10 s; (c) 7.4 cm/s
23. (a) 4.0 cm; 2.5 s; (b) no; $4.0\cos(2.5t)$ (cm)
25. 2.4 s
27. 1.4 s
29. —
31. (a) 1.0×10^{-2} rad/s; (b) 2.8 s; (c) 4.5×10^{-3} rad
33. —
35. —
37. $2\pi[l/(g-a_0)]^{1/2}$; the pendulum ceases to function
39. (a) $x'(t)=1.0\times10^{-3}\cos(32t)$ (m) (b) $x(t)=0.50t^2+1.0\times10^{-3}\cos(32t)$ (m)

Chapter 7

1. 30 J; −30 J
3. (a) 1.6×10^2 N; (b) 8.1×10^2 J
5. (a) 1.7×10^5 N; (b) -2.6×10^6 J; (c) 2.9×10^6 J
7. —
9. (a) 1.8; (b) 1.1
11. −3; 4/3
13. 430 J
15. 2.67 m
17. 1.8×10^{11} J
19. (a) $0.45(t^2+1)^2$ (J); (b) $6t$ (m/s^2); $0.6t$ (N); (c) 43 J
21. (a) -1.7×10^4 J; (b) 1.7×10^5 N
23. (a) $mgl\sin\alpha$; (b) $-mv_0^2/2$; (c) $-m[gl\sin\alpha+v_0^2/2]$
25. (a) 5.9×10^{14} J; (b) 1.1×10^4 m/s
27. (a) $-3mv_0^2/8$; (b) $3mv_0^2/16\pi R$
29. 441 W
31. (a) 0.20 m; (b) 20 N; (c) $(200t-40)$ (W)
33. —
35. —
37. $W=\frac{1}{2}m(v_2'^2-v_1'^2)+m\mathbf{a}_0\cdot(\mathbf{r}_2'-\mathbf{r}_1')$

Chapter 8

1. 1.03×10^4 J; -1.03×10^4 J
3. −0.15 J
5. 3.92×10^{10} J
7. (a) 4.5×10^{-3} J; (b) -6.0×10^{-3} J (c) -2.4×10^{-3} J
9. (a) $(-\alpha x^3/3+\beta x^4/4)$; (b) $F_0e^{-\beta x}/\beta$; (c) $-(F_0/\beta)\sin\beta x$
11. (a) 1.96 J; (b) 0.98 J; (c) 1.96 J
13. —
15. (b) 2.8 m/s
17. (a) 2.2 m/s; (b) $[1.96(1-\cos\theta)+1]^{1/2}$ (m/s)
19. (a) $[2gR(1-\cos\theta)+v_0^2]^{1/2}$
21. (a) 2.8 m/s; (b) 80 cm
23. (a) $[2g(h-2a)+v_0^2]^{1/2}$; (b) mg; $[2g(h-a)+v_0^2]m/a$
25. —
27. —
29. $m[g(3\cos\theta-2\cos\theta_0)+v_0^2/l]$
31. —
33. $mg[3\cos\phi+2(l-l_1)/l_1-2l\cos\theta_0/l_1]$
35. (a) -9.7×10^{-3} J; (b) 1.5 m/s; (c) 137°
37. (a) 1 J; (b) no; (c) −2 J; yes; x_2, x_1
39. —
41. (a) 2.0 J; (b) 2.0 J; 2.75 J; 3.0 J; (c) 0.37 m; zero
43. (c) $-(\mu mg/k)+[\mu^2m^2g^2+kmv_0^2]^{1/2}/k$; the amplitude must be positive

Chapter 9

1. (a) 18 cm/s; (b) 28 cm/s
3. (a) −0.1 m/s; (b) 0.01 J (gain)
5. $mv_0^2/2$; $3mv_0^2/2$
7. 0.15 m/s
9. (a) 4.2×10^2 m/s; (b) 0.6 m/s
11. (a) $0.2\,m/M$ (m/s); (b) $0.02\,m(1-m/M)$ (J)
13. (a) 3 m/s; (b) 160 m/s

15. $(M+m)V/m \cos\alpha$
17. 39.6 N-s
19. −0.75 N-s; the same
21. $2mv \cos\theta$ (upward)
23. —
25. —
27. —
29. (a) 2.5×10^5 m/s along incident direction; (b) 2.5×10^5 m/s; (c) 5.0×10^{-17} J per proton
31. (a) 6.2×10^6 m/s; (b) 5.7×10^{-14} J
33. —
35. 14.5°
37. (a) −0.33 m/s; (b) 1.17 m/s; (c) 1.9 m/s
39. 67 km/hr (north)
41. —
43. (a) 2.25 m/s (left); (b) 4.5 N-s (left); (c) 4.5 m/s (left)
45. —

Chapter 10

1. $F_0[z\mathbf{j}-y\mathbf{k}]$
3. 0.20 m; 100 N
5. 1.7×10^{42} J-s
7. (a) 2.1×10^3 m/s; (b) 3.1×10^{11} J-s
9. $a=6.7\times10^6$ m; $\epsilon=0.019$
11. 1.2×10^5 J-s
13. —
15. $-(0.35\mathbf{i}+0.78\mathbf{j}+0.52\mathbf{k})$
17. —
19. (a) $(N-mg\cos\alpha)x$; (b) zero; (c) $N=mg\cos\alpha$
21. —
23. $v_2=0.73$ m/s; $v_1=1.4$ m/s
25. —
27. (a) $k/2r^2$; (b) $mv_0^2/2 = m(v_r^2+v_\theta^2)/2+k/2r^2$; $mv_0b=mv_\theta r$; (d) $v_\theta = v_0/[1+k/mv_0^2b^2]^{1/2}$
29. (b) $a_0=5.3\times10^{-11}$ m; $E_1=-2.2\times10^{-18}$ J; $E_2=-5.4\times10^{-19}$ J
31. —
33. $a^3/T^2=3.34\times10^{18}$ m^3/s^2
35. $E=-3.8\times10^{32}$ J; $L=9.0\times10^{38}$ J-s
37. (a) $2mv_0a$; (b) $\Delta E=3mv_0^2/2$ (no longer bound)
39. 6.3 cm
41. $D=(2\omega v_0^3\cos\lambda)/3g^2$ (west)

Chapter 11

1. $\bar{x}=29$ cm; $\bar{y}=22$ cm
3. 60 cm
5. $(2/3)l-\lambda_0 l^2/6M$
7. —
9. (a) 3.0 m/s^2; (b) same as (a)
11. $0.063t^2-1.5t$ (m); (to the right)
13. (a) $\mathbf{F}_1=6t\mathbf{j}$ (N); $\mathbf{F}_2=16\mathbf{i}-12t\mathbf{j}$ (N); (b) $(16/3)\mathbf{i}-2t\mathbf{j}$ (m/s^2); (c) $16\mathbf{i}-6t\mathbf{j}$ (N)
15. (b) $-\mathbf{j}(2mv_0^2/R_0)\sin(v_0t/R_0)$ (c) mv_0^2/R_0, radially inward; (d) same
17. (b) $m_2dv_2/dt = k[l-(x_2-x_1)]$
19. —
21. (a) zero; (c) $mv_0^2=mv^2+k(l-d)^2/2$
23. (a) $(Mmg\cos\alpha)/(M+m\sin^2\alpha)$; (b) $Mg(M+m)/(M+m\sin^2\alpha)$; vertically up
25. —
27. —
29. (a) 49 N; (b) zero; (c) 2 rad/s
31. $T=H=2.0\times10^3$ N; $V=1.2\times10^3$ N
33. 57 N (perpendicular to wall); 113 N (perpendicular to plane)
35. $F=71$ N; $H=71$ N; $V=250$ N
37. (a) 7.3×10^{22} kg; (b) 9.1×10^{-31} kg (c) 8.4×10^{-28} kg
39. —
41. —

Chapter 12

1. (a) 45 J-s; (b) 675 J
3. 9.8×10^{37} kg-m^2
5. 1.4×10^{42} J-s; 2.5×10^{36} J
7. (a) 4.0×10^{-3} kg-m^2; (b) 4.0×10^{-3} kg-m^2
9. —
11. —
13. —
15. (a) 7.5 J; (b) 12 rad/s^2; (c) 24 rad/s
17. (a) 15 rad/s; (b) 5.6×10^2 J; (c) 5.6×10^2 J
19. (a) 2.0×10^4 J; (b) 80 J; (c) -2.0×10^4 J
21. 4.0×10^3 J
23. (a) $\frac{1}{2}at^2$; (b) $\frac{1}{4}Ma^2t^2$; (c) $ma^2t^2/2$

25. (a) 1.9 s; (b) 0.91 m
27. $\left\{\left[\frac{2}{5}r^2+(l+r)^2\right]\Big/ l(l+r)\right\}^{1/2}$
29. $Mg[1+2(1-\cos\theta_0)/(1+I_c/MD^2)]$; vertically upward
31. —
33. $[F_0b(b+c)]/(I_c+Mb^2)$
35. (a) $E=\frac{1}{2}Mv^2+\frac{1}{4}Mb^2\omega^2-Mgx\sin\beta=0$
(b) $\left(\frac{4}{3}gl\sin\beta\right)^{1/2}$
37. (a) $Mgh=\frac{1}{2}Mv^2+\frac{1}{5}Mb^2\omega^2+Mg(h-x\sin\beta)$;
(b) $\omega=[10gh/7b^2]^{1/2}$
39. $\omega=15$ rad/s; 1.1×10^3J
41. —

Chapter 13

1. 1.7 kg/m^3; yes
3. (a), (b), (c) 2.2×10^{-2} m^3; (d) yes
5. $n=10^6$ atoms/m^3; $\rho=1.7\times10^{-21}$ kg/m^3
7. 3.3×10^{-9} m
9. (a) 2.7×10^2 J; (b) 6.0×10^3 N/m^2
11. (a) 520 m/s; (b) 10 atm; (c) 1.1×10^4 J
13. (a) 8.0×10^{-8} m; 2.0×10^{-10} s;
(b) 8.0×10^{-7} m; 2.0×10^{-9} s;
(c) 8.0 m; 0.02 s
15. (a) $n_i=2N_i/V$; $P_i=2mN_iv_{th}^2/3V$;
(b) $n=(N_1+N_2)/V$;
$P=mv_{th}^2(N_1+N_2)/3V$;
$E=mv_{th}^2(N_1+N_2)/2$
17. (a) $P_1=9.9\times10^4$ N/m^2;
$P_2=1.2\times10^4$ N/m^2;
(b) $E_1=3.7\times10^3$ J;
$E_2=4.5\times10^2$ J;
(c) $v_{th}=460$ m/s; $P=5.5\times10^4$ N/m^2
19. —
21. (a) 7.6×10^4 N/m^2; 0.98 kg/m^3;
(b) 5.9×10^4 N/m^2; 0.76 kg/m^3;
(c) 4.4×10^4 N/m^2; 0.56 kg/m^3
23. (a) zero; (b) $P_B(V_A-V_B)$;
(c) $P_A(V_A-V_B)$
25. —
27. (a) 2.6×10^2 J; (b) 1.6×10^3 J

Chapter 14

1. (a) 1.29 atm; 0.29 atm; (b) 15 m
3. —
5. (a) ρghA; (b) ρghA
7. —
9. —
11. 2.86×10^4 N
13. 3.6 m/s^2
15. 32°
17. (a) $g\sin\alpha$
19. —
21. (a) F_1/A_1
23. (a) 2.5×10^2 N; (b) 2.0 m
25. —
27. 1.2×10^{-4} m^3
29. (a) ρ/ρ_0; (b) ρ_0l^2gh
31. (a) $W-\rho l^3g$
33. 13 cm/s
35. $v_0r^2/(r+y\tan\alpha)^2$; πr^2v_0
37. (a) 5.4 m/s; (b) 5.4×10^{-3} m^3/s
39. —
41. (a) 1.9 m/s; (b) 5.3 m/s;
(c) 6.0×10^{-2} m^3/s
43. 64 m/s

Chapter 15

1. (a) 558.3°R; (b) 310.2 K; (c) 37.0°C
3. (a) 0°; (b) 575°; (c) none
5. (a) 1.3 moles;
(b) 2.6×10^{25} molecules/m^3;
(c) 4.6×10^3 J
7. (a) 5.6×10^3 J; (b) 5.6×10^3 J;
(c) 6.2×10^3 J; (d) 7.5×10^3 J
9. —
11. (a) slope decreases
13. (a) 30 liters; (b) 20°C; (c) 0.40 atm
15. (a) 5.2×10^{25}/m^3; (b) 2.2 atm;
(c) 1.2×10^3 J
17. (a) 9.4×10^4 N/m^2; (b) 291 K;
(c) $P_{He}=9.1\times10^4$ N/m^2;
$P_A=6.0\times10^4$ N/m^2;
$P_f=1.5\times10^5$ N/m^2
19. (a) $P_A=4.3\times10^4$ N/m^2;
$P_B=1.2\times10^4$ N/m^2;
(b) 4.8×10^{24}/m^3; (c) 280 K;
1.9×10^4 N/m^2

21. (a) 346 K; (b) 910 J
23. (a) 3.9×10^4 N/m^2; (b) 9.7×10^3 N/m^2; 2.2×10^3 J
25. 7.1×10^{-3} cm
27. —
29. (a) 2.5 cm^2; (b) 19 cm^3; (c) 6.6×10^{-3} cm
31. —
33. (a) 0.34 percent; (b) 0.17 percent; (c) 0.51 percent; (d) −0.51 percent
35. (a) $4\pi r_0^3 \alpha\ \Delta T$; (b) $4\pi r_0^3 \beta\ \Delta T/3$

Chapter 16

1. (a) 1.1×10^5 N/m^2; (b) −490 J; (c) 490 J
3. —
5. 160 J
7. (a) 0.25 m/s; 6.3×10^{-2} J; (b) 6.0 cal; (c) 0.10 K
9. —
11. $3\mu R(T_2 - T_1)/2 + P_3(V_3 - V_1)$
13. (a) $T_i = 3300$ K; $T_f = 2700$ K; (b) -9.0×10^4 J; (c) 2.9×10^4 J
15. (a) 1.12 atm; (b) 12 cm
17. (a) 2.2 moles; (b) 2.6×10^{-2} m^3; (c) $T_A = 1.2 \times 10^3$ K; $T_B = 430$ K; (d) 2.9×10^4 J
19. (a) $P_0 + 10R/V_0$; (b) $2[P_0 + 10R/V_0]$; (c) 125 J; (d) 250 J
21. —
23. —
25. 6.3°C
27. 6.2×10^5 cal
29. 41°C
31. 25 g
33. —
35. 2.4×10^6 cal/s
37. (a) 6.4×10^{-3} cal/s; (b) 8.0×10^{-5} cm/s
39. (b) $T(r) = \dfrac{T_2 \ln(R_1/r) + T_1 \ln(r/R_2)}{\ln(R_1/R_2)}$

Chapter 17

1. −290 cal/K; +290 cal/K
3. 1.1 cal/K
5. (a) −62 cal/K; (b) 69 cal/K
7. (a) 86°C; (b) 0.79 cal/K
9. (a) 18 cal/K; (b) 17 cal/K; (c) 24 cal/K
11. —
13. (a) $R \ln \alpha$; (b) $-R \ln \alpha$
15. (a) zero
17. (a) 25°C; (b) 12 J/K
19. —
21. —
23. 19 percent; 650 J
25. (a) 17 percent; (b) $Q_1 = 1.2 \times 10^4$ cal; $W = 8.4 \times 10^3$ J
27. 24 cal
29. (a) 606 K; (b) 150 cal

Chapter 18

1. 71 m/s
3. —
5. (a) 0.96 m/s; (b) 0.094 m
7. (a) 537 m/s; (b) 46 K
9. (a) $a^3/[a^2 + (x - ut)^2]$; (b) $a^3/[a^2 + (x + ut)^2]$
11. —
13. $y = 0.02 \cos[(10\pi/3)(x - 60t)]$ (m)
15. (a) 50 cm; 20 Hz; (b) 377 cm/s; (c) 4.7×10^4 cm/s^2
17. (a) 7.1 Hz; (b) 2800
19. (a) 410 m/s; (b) 410 Hz; (c) 820 Hz
21. 1.6
23. (a) 825 Hz; (b) 825 Hz; (c) 413 Hz
25. 495 Hz
27. (a) 4 beats/s; (b) 2.73 beats/s
29. (a) 461 Hz; (b) 339 Hz
31. —
33. 350 m/s
35. 4.4 J/m
37. —

Solutions to Odd-Numbered Problems

Chapter 19

1. (a) 9.63×10^4 C;
   (b) -9.63×10^4 C
3. 58 N; repulsive
5. 1.2×10^2 N; attractive
7. 1.44×10^{-37} C
9. $F_y = \frac{qQ}{4\pi\epsilon_0} y[x^2 + y^2 + z^2]^{-3/2}$;
   $F_z = \frac{z}{y} F_y$
11. (a) 4.8×10^{-2} N; (b) 6.2×10^{-2} N;
    (c) 6.5×10^{-2} N
13. (a) $0.91\, q^2/4\pi\epsilon_0 a^2$; directed toward center of square;
    (b) same as (a)
15. (a) zero; (b) $ybQq/\pi\epsilon_0[b^2 - y^2]^2$
17. $Qqa/2\pi\epsilon_0[a^2 + y^2 + z^2]^{3/2}$
19. along the axis and away from the circle if $qQ > 0$
21. (a) $dF_H = Q\lambda a^2 d\theta/4\pi\epsilon_0[a^2 + b^2]^{3/2}$;
    $dF_V = bdF_H/a$;
    (b) $Q\lambda ab/2\epsilon_0[a^2 + b^2]^{3/2}$; directed vertically up for $Q\lambda > 0$

Chapter 20

1. 4.5×10^8 N/C; 4.5×10^6 N/C; 4.5×10^2 N/C; directed radially outward
3. (a) zero; (b) 2.4×10^4 N/C in the $-x$ direction; (c) 2.8×10^4 N/C in the $+x$ direction
5. (a) $p/9\pi\epsilon_0 a^3$ in $+x$ direction;
   (b) same as (a)
7. —
9. 1.8×10^{-9} C/m^2
11. (a) 1.0×10^{-7} N/C; (b) 9.4×10^5 m
13. $\frac{\lambda}{2\pi\epsilon_0}\left[\frac{1}{y} - \frac{1}{(y^2 + l^2)^{1/2}}\right]$ parallel to line charge, toward $-\lambda$ part
15. 9.5×10^{-11} C/m
17. $\lambda/2\pi\epsilon_0 a$, perpendicular to the diameter joining the endpoints
19. (a) zero
21. $\frac{\sigma b}{2\epsilon_0}\left[\frac{1}{(a^2 + b^2)^{1/2}} - \frac{1}{(R_0^2 + b^2)^{1/2}}\right]$; along axis
23. —

25. **(b)** $\frac{q}{2\epsilon_0}\left[1-\frac{a}{(R_0^2+a^2)^{1/2}}\right]$;
(c) $q/2\epsilon_0$
27. **(a)** zero; **(b)** $-2.5\ \mu$C; **(c)** $+2.5\ \mu$C
29. **(a)** $-r\rho_0/2\epsilon_0$; **(b)** $-\rho_0 a^2/2\epsilon_0 r$;
(c) zero
31. **(b)** $\rho_0 r(a^2-r^2)/4\epsilon_0 a^2$; radially outward
33. —
35. **(a)** $Ze/4\pi\epsilon_0 r^2$;
(b) $7.5\times10^{44} Zer/4\pi\epsilon_0 A$
37. **(a)** $\rho_0 r/3\epsilon_0$; **(b)** $\rho_0 a^3/3\epsilon_0 r^2$;
(c) $\rho_0(a^3+b^3-r^3)/3\epsilon_0 r^2$;
(d) $\rho_0(a^3+b^3-c^3)/3\epsilon_0 r^2$
39. —
41. **(a)** $-q/4\pi a^2$; $q/4\pi b^2$;
(b) $-q/4\pi a^2$; zero
43. $q/4\pi\epsilon_0 r^2$ for $r\leqslant a$;
zero for $a\leqslant r\leqslant b$;
$(Q+q)/4\pi\epsilon_0 r^2$ for $r\geqslant b$

Chapter 21

1. **(a)** -6.0×10^{-4} J; **(b)** 6.0×10^{-4} J;
(c) zero; **(d)** -5.2×10^{-4} J
3. **(a)** 5.7×10^4 V/m; **(b)** planes parallel to the charged plane;
(c) -1.1×10^4 V
5. **(a)** 1.0×10^4 V/m; **(b)** -6.0×10^{-4} J
7. **(a)** $\frac{q}{4\pi\epsilon_0}\left[\frac{3d-4a}{a(d-a)}\right]$; **(b)** $3d/4$
9. **(a)** zero; **(b)** zero; **(c)** zero
11. 2.3×10^{-2} J; -1.3×10^{-2} J;
1.1×10^{-2} J
13. **(a)** $\frac{q}{\pi\epsilon_0}(y^2+a^2/2)^{-1/2}$;
(b) $\frac{qy}{\pi\epsilon_0}(y^2+a^2/2)^{-3/2}$; directed vertically away from square
15. $2\alpha(x\mathbf{i}+y\mathbf{j}-2z\mathbf{k})$
17. —
19. **(a)** zero;
(b) $\frac{q_0\sigma}{2\epsilon_0}[R-(R^2+z^2)^{1/2}+\sqrt{z^2}]$
21. —
23. **(a)** $\frac{Ze}{4\pi\epsilon_0 r}$; **(b)** $6.8\times$
$10^{54}\frac{Ze}{A}(1.8\times10^{-30}A^{2/3}-r^2/2)$
(meters); **(c)** 3.4×10^{-4} J
25. **(b)** $-\sigma a/\epsilon_0$; **(c)** $-\sigma(d-b)/\epsilon_0$
27. **(b)** $\frac{Q_0}{4\pi\epsilon_0 a}$
29. $\frac{q^2}{4\pi\epsilon_0 a}[4+\sqrt{2}]$
31. -4.4×10^{-18} J
33. $\frac{-3q^2}{2\pi\epsilon_0 a}$; $\frac{3q^2}{4\pi\epsilon_0 a}$
35. **(b)** 2.9×10^{21} V/m
37. 4.6×10^{-13} J; 1.5×10^{-10} J
39. **(a)** -1.1×10^{-3} V; **(b)** 26 V
41. **(a)** -4.7×10^{-2} V; **(b)** 2.4×10^2 V/m
43. **(a)** $\frac{\rho_0}{3\epsilon_0}\left[\frac{a^2}{b}(b-a)+\frac{(a^2-r^2)}{2}\right]$
(b) $\frac{\rho_0 a^3}{3\epsilon_0}\left(\frac{1}{r}-\frac{1}{b}\right)$; **(c)** $-\frac{\rho_0 a^3}{3b^2}$
45. $\frac{\rho_0}{2\epsilon_0}(b^2-a^2)$
47. —

Chapter 22

1. 1.2×10^{-5} C
3. **(a)** $1.2\times10^{-5}\ \mu$F;
(b) 1.2×10^{-7} C/m^2
5. **(a)** $1.1\times10^{-3}\ \mu$F; **(b)** $1.1\times10^{-2}\ \mu$C;
(c) $8.8\times10^{-2}\ \mu$C/m^2;
$-8.9\times10^{-2}\ \mu$C/m^2
7. **(a)** 1.3×10^4 V; **(b)** 6.3×10^6 V/m
9. 0.078F
11. **(a)** 8.0 μF; **(b)** 1.9 μF
13. 1.6 μF
15. 3.2 μF
17. **(a)** zero on C_5, 15 μC on each of the others; **(b)** 3.0 μF
19. **(a)** 6 μF; 1.5 μF; **(b)** 1 μF; 2 μF;
4.5 μF; 9 μF;
(c) 12 μF; 0.75 μF; 1.2 μF; 3.0 μF;
2.3 μF; 4.0 μF; 5.0 μF; 1.8 μF;
7.5 μF
21. **(a)** 5.6×10^{-5} J; 1.9×10^{-5} J;
(b) 7.5 μC; 23 μC; **(c)** 5.6×10^{-5} J
23. $\frac{\epsilon_0 A\kappa_1\kappa_2}{a\kappa_2+b\kappa_1}$
25. —
27. —
29. **(b)** $\pm Q(\kappa-1)/\kappa$; **(c)** no
31. $4\pi\epsilon_0\left[\frac{1}{a}-\frac{1}{b}+\frac{(1-\kappa)}{\kappa}\left(\frac{1}{c}-\frac{1}{d}\right)\right]^{-1}$;
yes

33. —

35. (a) $-\dfrac{q}{4\pi a^2}\left(\dfrac{\kappa-1}{\kappa}\right)$; (b) $\dfrac{q}{4\pi b^2}\left(\dfrac{\kappa-1}{\kappa}\right)$

37. (c) $-\epsilon_0 A V_0^2 dy/y^2$

39. (b) $\dfrac{\epsilon_0 V^2}{2d}[A + ay(\kappa-1)]$

41. —

43. —

Chaper 23

1. (a) 0.20 A; (b) 4.0×10^4 A/m^2

3. 2.5×10^5 A/m^2; 1.3×10^5 A/m^2

5. (a) 3.3 A; (b) 200 C; (c) 1.9×10^{22} electrons

7. $i_{Ag}/i_{Al} = 1.8$

9. 160 Ω

11. (a) 100 Ω; (b) 1.0 A

13. (a) $10\,i$; (b) $50\,i$

15. (a) 0.40 Ω; (b) 4.0×10^5 Ω

17. (a) 2.5 μs; (b) 4.0 A; (c) 1.0×10^{-3} J; (d) $800\exp\{-2t/2.5\,\mu\text{s}\}$ (W)

19. (a) 0.50 A; (b) 91 μC; 460 μC; (c) 0.41 A; 0.041 A

21. (a) $5\times10^{-4}[1-\exp(-10^3 t)]$ (C) (b) $0.5\exp(-10^3 t)$ (A) (c) 2.5×10^{-2} J

23. (b) $C\varepsilon^2/2$

25. $\dfrac{\varepsilon}{d}(1-e^{-t/RC})$; $C = \varepsilon_0 A/d$

27. (c) $\dfrac{Q_0}{RC_1}\exp[-t(C_1+C_2)/RC_1C_2]$

29. (a) 2.6×10^{-8} Ω-m; (b) 3.7×10^{-8} Ω-m

31. —

33. 677 Ω

Chapter 24

1. (a) 23 Ω; (b) 1.7 Ω

3. 2R/3; 1:1:2

5. —

7. (a) 240 Ω; (b) 0.50 A; (c) 60 W

9. 0.68

11. —

13. —

15. (b) 4.6 W by 10-volt battery; 12 W by 20-volt battery

17. (a) 0.50 A; (b) $i_{25} = 1.4$ A (up); $i_{50} = 1.9$ A (right); $i_{10} = 0.11$ A (left); (c) no

19. —

21. (a) 1.0×10^{-5} s; (b) 30 A; (c) 300 μC

23. (a) 5.0 A; (b) zero on the 3 μF capacitor; 100 μC on the 2 μF capacitor; zero

25. (a) 3.8 A; (b) same as Problem 23b

27. (a) 0.050 A (left); (b) 0.30 A (counterclockwise); (c) 0.35 A (clockwise).

29. ε/R_1 through R_1; $\varepsilon/2R_2$ through R_2; $\varepsilon/2R_3$ through R_3

31. (b) $i_1 = i_3 = 5.0$ A; $i_2 = 0$

33. —

Chapter 25

1. (a) 1.5×10^{-3} N (west); (b) zero; (c) 1.1×10^{-3} N (vertically up)

3. (a) 1.1×10^{-16} N (east); (b) 1.2×10^{14} m/s^2

5. positive z-axis

7. (a) zero; (b) $\mu_0 idl/4\pi a^2$ (along $-x$ axis); (c) $\mu_0 idl/4\pi a^2$ (along $-y$ axis) (d) $\mu_0 idl/8\pi a^2$ (in x-y plane midway between the y and $-x$ axis)

9. —

11. —

13. 3.6×10^{-5} T

15. —

17. $\dfrac{\mu_0 i}{2\pi}\left\{\left[\dfrac{y-2a}{x^2+(y-2a)^2}-\dfrac{y}{x^2+y^2}\right]\mathbf{i} + \left[\dfrac{x}{x^2+y^2}-\dfrac{x}{x^2+(y-2a)^2}\right]\mathbf{j}\right\}$

19. 40 A; 1.6×10^6 A

21. $\mu_0 i\alpha/4\pi a$ (down)

23. $\mu_0 i/4a$

25. —

27. —

29. 1.6×10^3 turns/m

31. zero

33. —

35. (a) $-\mu_0 Ni$; (b) $+\mu_0 Nih/l$ (c) zero; (d) zero

37. (a) $\mu_0 iR/(a^2+R^2)^{1/2}$

39. (a) $\mu_0 ir/2\pi a^2$; (b) $\mu_0 i/2\pi r$; (c) $\dfrac{\mu_0 i}{2\pi r}\left[\dfrac{c^2-r^2}{c^2-b^2}\right]$; (d) zero

Chapter 26

1. (a) 210 m; (b) 1.3×10^{-3} s; (c) 4.8×10^{3} rad/s
3. (a) $\frac{R_\alpha}{R_p} = 2$; (b) $\frac{\omega_\alpha}{\omega_p} = \frac{1}{2}$; (c) $\frac{E_\alpha}{E_p} = 4$
5. 12 cm below slit; $E(^7\text{Li}):E(^6\text{Li}) = 7:6$
7. (a) 7.2×10^{7} rad/s; (b) 69 orbits; (c) 3.0×10^{-6} s
9. 1.2×10^{4} MeV; no
11. (a) $m\frac{d\mathbf{v}}{dt} = q\epsilon\frac{\mathbf{v}\times\mathbf{r}}{r^3}$
13. (a) 9.6×10^{16} rad/s (b) 4.6×10^{-10} m
15. —
17. (a) iaB directed perpendicular to the wire and in the plane of the coil
19. —
21. 9.7×10^{-5} N between longer wires; 2.4×10^{-6} N between shorter wires
23. $\frac{\mu_0 Iic}{2\pi}\left[\frac{1}{a} - \frac{1}{b}\right]$; (left); yes
25. —
27. (a) 0.63 A-m^2; (b) 4.0×10^{-2} A-m^2; (c) 0.31 A-m^2
29. 9.4×10^{-3} N-m
31. —
33. $il^2B/4\pi$
35. (a) 1.3×10^{-5} m/s; (b) 2.5×10^{-10} m^3/C (c) 2.5×10^{28} charge carriers/m^3

Chapter 27

1. (a) 2.0 V/m; along $+y$ direction (b) zero; (c) 1.0 V/m along $+z$ direction
3. (a) 1.6×10^{-5} T (perpendicular to plane of wires) (b) 4.8×10^{-4} V/m (perpendicular to the wires and in their plane); (c) 3.2×10^{-4} V/m; (along the wires)
5. $\frac{\mu_0 ivd}{2\pi x(d-x)}$; directed to the right
7. $(\mu_0\rho_0 a^3/3r^3)\mathbf{v}\times\mathbf{r}$ (outside) $(\mu_0\rho_0/3)\mathbf{v}\times\mathbf{r}$ (inside)
9. (a) 1.0 V/m (up); (b) 0.20 V (upper end)
11. 5.6×10^{-4} V
13. (a) 17 Ω; (b) 3.8×10^{-3} W; (c) 7.5×10^{-4} N; (d) 3.8×10^{-3} W
15. (b) $(Blv/R)\cos\alpha$; clockwise as viewed from above
17. (b) $\omega Bl^2/2$; end at P at higher potential
19. —
21. $\frac{\mu_0 iv}{2\pi}\ln\left[1 + \frac{l}{a}\cos\alpha\right]$
23. (a) zero; (b) $\frac{1}{2}\omega Bl^2\sin\omega t$ (upward in one, down in the other); (c) $\omega l^2B\sin\omega t$
25. —
27. (a) counterclockwise; (b) 6.3×10^{-4} V; (c) 6.3 μA
29. (a) $\mu_0 n\pi a^2(i_0 + \alpha t)$; (b) $\mu_0 n\pi a^2\alpha$; (c) $\mu_0 n\pi a^2\alpha/R$; opposite to solenoid current
31. (a) $(\mu_0 l\alpha/2\pi)\ln(1 + b/a)$; clockwise (b) same as (a)
33. (a) $\mu_0 i_0(b - \sqrt{b^2 - a^2})\sin\omega t$ (b) $\omega\mu_0 i_0(b - \sqrt{b^2 - a^2})\cos\omega t$
35. —
37. (a) $q\alpha r/2$; directed to the right and tangent to a concentric circle; (b) 8.8×10^{8} m/s^2; opposite to the direction in (a)
39. $m\frac{d\mathbf{v}}{dt} = \frac{q}{2}\mathbf{r}\times\frac{d\mathbf{B}}{dt} + q\mathbf{v}\times\mathbf{B}$

Chapter 28

1. (a) 6.3×10^{-3} H/m; (b) 2.4×10^{-6} Wb; (c) 4.8×10^{-3} Wb
3. (a) 66 μH; (b) 6.6×10^{-7} Wb
5. —
7. (a) 10 mH; (b) 5.0×10^{-6} Wb; 0.10 Wb
9. $\frac{\mu_0 li}{\pi}\ln\left(\frac{d-a}{a}\right)$
11. (a) 1.0×10^{-4} s; (b) 6.3×10^{-2} A; (c) 10 V
13. —
15. —

17. —
19. —
21. —
23. **(a)** $ABCDEFA$; **(b)** $ABEFA$
25. **(a)** 1.0 A; **(b)** 0.83 A; **(c)** 42 μC
27. **(a)** 0.50 A; **(b)** 0.67 A; **(c)** 4.0 A; **(d)** 4.7 A
29. **(a)** 1.6×10^4 rad/s; **(b)** 4.0×10^{-4} s; **(c)** 6.3×10^{-6} J
31. —
33. —
35. —
37. —
39. —

Chapter 29

1. 5.5×10^{14} Hz; $1.1 \times 10^7\ \text{m}^{-1}$
3. **(a)** 5.0×10^{14} Hz; **(b)** 3.1×10^{15} rad/s; **(c)** $1.0 \times 10^7\ \text{m}^{-1}$
5. **(a)** 95 V/m; **(b)** 3.2×10^{-7} T; **(c)** no
7. Saturn; 16W; Pluto; 0.86 W
9. 4.1×10^{-7} m; 5.5×10^{-7} m
11. **(b)** $\dfrac{\epsilon_0 \mathscr{E}}{RCd}\, e^{-t/RC}$
13. yes
15. —
17. —
19. $7.5 \times 10^8\ \text{m}^2$
21. zero; 46°
23. —
25. 1.2×10^8 m/s; 2.9×10^{-7} m
27. 3.7×10^{-7} m
29. —
31. $\dfrac{2q^2 a_0^2}{3\pi\epsilon_0 c^3}$
33. **(a)** qvB/m
35. 7.1×10^{-24} J

Chapter 30

1. **(a)** 500 s; **(b)** 3.1×10^3 s; 2.1×10^3 s
3. 1.5; 3000 Å
5. 2α
7. —
9. **(a)** 42°; **(b)** 70°; **(c)** no emerging beam
11. **(a)** 26°; **(b)** 78°
13. 1.53
15. **(a)** 16°; 19°; **(b)** 3.2°
17. $\sqrt{2}$
19. 31°
21. —
23. 96%; reflected
25. 5.8
27. —
29. —
31. —

Chapter 31

1. 90 cm
3. **(a)** 30 cm behind mirror; **(b)** 3.0; **(c)** virtual and erect
5. **(a)** 0.75 m in front of mirror; **(b)** -7.5×10^{-6}; **(c)** 1.5×10^{-4} m
7. **(a)** 13 cm behind mirror; virtual **(b)** 0.67; **(c)** 6.7 cm; erect
9. **(a)** $D = s^2/(s + R/2)$; **(b)** zero
11. —
13. —
15. **(a)** 18 cm behind mirror; 3.0 cm high; erect; **(b)** 15 cm behind mirror; 3.8 cm high; erect
17. -1
19. **(a)** at center of bowl; **(b)** erect
21. a distance $\dfrac{(d+h)(4d+3h)}{h}$ above mirror
23. **(a)** 8.8 R; virtual image in water **(b)** 8.8 R; real image in glass
25. **(a)** 90 cm to right of left surface **(b)** real; inverted; 4.0 cm; **(c)** 10 cm to right of right surface; **(d)** inverted; real; -1
27. 1.6 R to left of left surface of the sphere
29. **(a)** 20 cm; **(b)** -20 cm; **(c)** converging
31. —
33. **(a)** 60 cm; **(b)** 20 cm
35. —
37. **(a)** away from lens; **(b)** 12 cm
39. —
41. 10 cm to left of the second lens; virtual; inverted; 20 cm high
43. 12 cm to the right

45. —
47. —
49. —

Chapter 32

1. 3.9 V/m
3. **(a)** 1.2π rad, or 0.23π rad; **(b)** 2.0×10^{-7} m
5. 6.8×10^{-5} m
7. 64 cm
9. —
11. $\Delta = m\lambda$ for a maximum; $\Delta=(m+1/2)\lambda$ for a minimum; $m=0,1,2,\ldots$
13. —
15. —
17. —
19. **(a)** 5.0×10^{-4} cm; **(b)** 7.6°; 15°; **(c)** 7
21. 2.9×10^{-4} cm
23. $\dfrac{13{,}000}{m}$ Å; $m=1,2,3,\ldots,$
25. —
27. —
29. —
31. —
33. **(b)** $I_m(\sin^2\beta/\beta^2)$
35. 6.4×10^{4} m
37. 1.3 cm
39. $\dfrac{3.0\times10^{4}}{m+\frac{1}{2}}$ Å; $m=0,1,2,\ldots$
41. **(a)** 107; **(b)** 0.94 mm
43. 1.8×10^{-5} cm; 10,000 Å
45. **(a)** 1.4 m; **(b)** 4.3 mm
47. —

Chapter 33

1. 59°
3. —
5. **(a)** 0.1 W/m^2; **(b)** 0.075 W/m^2
7. **(a)** linearly polarized; **(b)** linearly polarized; **(c)** left circularly polarized; **(d)** elliptically polarized; **(e)** elliptically polarized
9. **(a)** 1/8; **(b)** zero
11. —
13. 1
15. **(a)** 1.86×10^{8} m/s; **(b)** 1.94×10^{8} m/s;
17. **(a)** $0.67c$; $0.60c$; **(b)** $0.65c$; $0.64c$
19. 5.2×10^{-3} rad
21. **(a)** A is e-ray; B is o-ray **(b)** A (B) is linearly polarized perpendicular to (in) plane of diagram; **(c)** 5.6°
23. **(a)** $d_0=\dfrac{\lambda}{2|n_e-n_o|}$; **(b)** $3d_0$; $5d_0$; ... **(c)** 17,000 Å
25. **(b)** at angle θ to the vertical
27. —
29. $\cos^2 2\theta$
31. —
33. —

Chapter 34

1. **(a)** 9.0×10^{13} W/m^3; **(b)** 1.8×10^{5} W/m^3; **(c)** 8.1×10^{-20} W/m^3
3. 530 W/m^2
5. —
7. 1.1×10^{-3} m; yes
9. **(a)** 7.2×10^{14} Hz; **(b)** 1.1 eV
11. 1.1×10^{45} photons/s
13. **(a)** 6.2×10^{-12} m; **(b)** 8.6×10^{-12} m; **(c)** 1.7×10^{5} eV
15. 4.56×10^{14} Hz; 6.58×10^{-7} m
17. —
19. n^2a_0
21. 1.7×10^{-36} m; no; λ is too small

Index